PROBLEMS IN MATHEMATICAL STATISTICS
Part I

Maurizio Tiso

University of Minnesota
School of Statistics
Technical Report # 638
December 15, 2000

PROBLEMS IN MATHEMATICAL STATISTICS
Part I

A Workbook with Solutions

Maurizio Tiso

PREFACE

Mathematical Statistics is in my opinion a misleading name for the branch of Statistics dealing with the mathematical aspects of the discipline. I find the label *Mathematical Methods of Statistics* more suitable to describe its contents and purposes. Although I decided to stick with the traditional name of the subject, I have nevertheless organized the problems in chapters according to the main mathematical tool required in their solutions; e.g. probability inequalities, limit theorems, characteristic functions, and so on. This is somewhat arbitrary since, for example, it is not clear if a problem that asks the student to prove the consistency of an estimator for a linear regression model should appear under the limit theorems or the linear model sections. In fact, a linear model is technically speaking, a mathematical tool of statistics as much as stochastic limit theory. Despite this, I am positive that this collection will prove beneficial to students and in particular to those preparing to take their Ph.D. qualifying exams as it should reveal very quickly areas of strenght or weakness.

Perhaps it is superfluous to say it, but interesting problems in Mathematical Statistics are necessarily non trivial and I realized at an early stage of my graduate school days that I was often working on the same problems again and again. This is inevitable as over periods of months or years one is likely to forget the steps that were instrumental in proving specific results. This was not efficient and I began to write down both the problems and their solutions. And this should explain how the idea of collecting problems was born. Only a subset of the entire collection is presented here, but I hope to be able to review more problems in the near future and have them ready for students to use.

The original inputs were problems assigned during the classes in Mathematical Statistics I took from Proff. Pruitt and Tierney in 1995, problems from old Ph.d. exam files available at the School of Statistics of the University of Minnesota and the classes in Probability taught also in 1995 by Proff. Gray and Fristedt at the School of Mathematics. Later on, I added problems I encountered in my research and a few were taken from textbooks I read. I tried to acknowledge the source of a problem when the attribution could be done without doubts. It is surprising how common some of the problems are or how often some problems presented in textbooks can be found in similar books written ten or twenty years earlier. If anybody is aware that any of the problem has a specific source or that some attributions are not correct, I would appreciate if they could let me know about it.

The solution are extremely detailed and some may say too detailed. This is not done to insult the readers' intelligence; it is simply the realization that computations can also be tricky at times and I did not want to find myself working on the same algebraic manipulations over and over again. I hope that this will be beneficial also to those students who are approaching the subject on their own or do not have easy access to Statistics faculty members.

I put a huge number of hours on working out the problems and at times I could not avoid asking myself if it was worth the time and effort. Today, a few years later, I feel I can say that it was worth it. Besides, now that the graduate

school days are over, when I browse through the problems, I see summations, integrals, laws of large numbers, central limit theorems, but I also see a few guys who shared many a day (and night) working on some assignment. Then I think of Alexandre Varbanov and our challenges with an old motorcycle videogame in between solving problems and his wife Milena who baked delicious cookies; I think of Grant Runyan and the dinners we shared to console ourselves of problems that did not want to be won: quite a few pounds we both gained were the direct consequence of the frustration with some of the problems presented here; and I think of Garrick Wallstrom who did not work with me at solving assignments, but still helped suggesting many interesting problems together with quite a few videogames. We also carried a huge couch in the rain for almost a mile. But that, I swear, had nothing to do with solving problems in Statistics. Finally, I must thank other fellow students, whose pictures do not appear anymore on the walls of the School, but are still in my heart: Ming-Dauh Wang, Pawel Stryszak, George Vesely, ... and Panagiotis Tsiamyrtzis (Panagiotis confess that for a moment you thought I was going to leave you out!) and many others. I must also acknowledge the help of many faculty members most of whom were always available to answer a question or clarify an issue. In particular, I must single out the help provided by Charlie Geyer, Morris Eaton, Ron Pruitt, Bert Fristedt and Bill Sudderth. It is inevitable that, despite my care, some errors still exist. As much as I would like to be able to blame them on somebody else, I believe that I have to take responsability for those. I would however appreciate if readers would use me the kindness to let me know about them so that I can amend the same. I am also interested in learning about different and more efficient solutions to some of the problems presented here.

Maurizio Tiso

St. Paul, Minnesota November 2000

Chapter 1

MOMENTS AND INEQUALITIES

1.1 Technicalities: Integration by Parts, Fubini's Theorem

PROBLEM 1.1.1. Let X be a random variable with $E[| X |] < \infty$.

(a) Show that

$$E[X] = \int_0^\infty (1 - F(x))dx - \int_{-\infty}^0 F(x)dx.$$

(b) If $X > 0$, find a, b, and c, which may all depend on t, in the following formula:

$$E[X - t \mid X > t] = \int_a^b c(1 - F(x))dx$$

with $t \geq 0$.

SOLUTION. There are two techniques to prove the statement of the problem: integration by parts and Fubini's Theorem.

Method 1: integration by parts. Since, by assumption $E[| X |] < \infty$, we have that $E[X]$ exists and can be written as

$$E[X] = E[X^+] - E[X^-]$$

where $X^+ = X \vee 0$ and $X^- = -(X \wedge 0)$ are the positive and the negative parts of X, respectively. The problem of computing $E[X]$ is thus reduced to the computation of $E[X^+]$ and $E[X^-]$. To do this we use the next two propositions:

Proposition 1. *If X is a nonnegative r.v., then* $E[X] = \int_0^\infty [1 - F(x)]dx$.

Proof. Using integration by parts yields

$$\begin{aligned}
E[X] &= \int_0^\infty x dF(x) = \lim_{y\to\infty} \int_0^y x dF(x) \\
&= \lim_{y\to\infty} \left[xF(x) \mid_0^y - \int_0^y F(x)dx \right] \\
&= \lim_{y\to\infty} \left[yF(y) - 0F(0) - \int_0^y F(x)dx \right] \\
&= F(\infty) \lim_{y\to\infty} \int_0^y dx - \lim_{y\to\infty} \int_0^y F(x)dx \\
&= \lim_{y\to\infty} \int_0^y (1 - F(x))dx = \int_0^\infty (1 - F(x))dx.
\end{aligned}$$

□

Proposition 2. *If X is an $\mathbf{R}^-$-valued random variable, then*

$$E[X] = \int_{-\infty}^0 -F(x)dx = \int_0^\infty -F(-x)dx.$$

Proof. The argument of the proof is again based on integration by parts. In fact,

$$E[X] = \int_{-\infty}^0 x dF(x) = \lim_{y\to-\infty} \int_y^0 x dF(x)$$

$$\begin{aligned}
&= \lim_{y\to-\infty} \left[xF(x) \mid_y^0 - \int_y^0 F(x)dx \right] \\
&= 0F(0) + \lim_{y\to-\infty} -yF(y) - \lim_{y\to-\infty} \int_y^0 F(x)dx.
\end{aligned}$$

As $y \leq 0$, we must have $-yF(y) \geq 0$ and therefore

$$0 \leq -yF(y) = -y \int_{-\infty}^y dF(x) \leq \int_{-\infty}^y x dF(x) \to 0$$

when $y \to -\infty$. This yields $E[X] = \int_{-\infty}^0 -F(x)dx$ or, after performing the change of variable $y = -x$, $E[X] = \int_0^\infty -F(-x)dx$. □

Now, if we combine Propositions 1 and 2, we get

$$E[X] = E[X^+] - E[X^-] = E[X^+] + E[-X^-];$$

and, since $X^+ \geq 0$ while $-X^- \leq 0$,

$$E[X] = \int_0^\infty [1 - F(x)]dx + \int_{-\infty}^0 -F(x)dx = \int_0^\infty [1 - F(x)]dx - \int_{-\infty}^0 F(x)dx^1.$$

[1]This tells us that $E[X]$ is finite iff both the integrals in this formula are finite.

For the second part[2], let $Y = X - t \mid X > t$. As by assumption $t \geq 0$, it follows that Y is an $\mathbf{R}^+$-valued r.v. and

$$\begin{aligned} G_Y(y) = P[Y \leq y] &= P[X - t \leq y \mid X > t] = P[X \leq y + t \mid X > t] \\ &= P[t < X \leq y + t]/P[X > t] = \frac{F(y+t) - F(t)}{1 - F(t)} \end{aligned}$$

Using Proposition 1 it is possible to write

$$\begin{aligned} E[X - t \mid X > t] = E[Y] &= \int_0^\infty (1 - G_Y(y))dy \\ &= \int_0^\infty \Big[1 - \frac{F(y+t) - F(t)}{1 - F(t)}\Big]dy \\ &= \int_0^\infty \Big[\frac{1 - F(y+t)}{1 - F(t)}\Big]dy \\ &= \int_0^\infty \frac{1}{1 - F(t)}\Big[1 - F(y+t)\Big]dy \\ &= \int_t^\infty \frac{1}{1 - F(t)}[1 - F(x)]dx. \end{aligned}$$

Hence, we showed that $E[X - t \mid X > t] = \int_a^b c(1 - F(x))dx$ with: $a = t$, $b = \infty$, and $c = (1 - F(t))^{-1}$.

Method 2: Fubini's Theorem. Writing X as the sum of its positive and negative parts, we have

$$E[X] = E[X^+] - E[X^-]$$

which can also be written as

$$\begin{aligned} E[X] &= \int_0^\infty x dF(x) - \int_{-\infty}^0 -x dF(x) \\ &= \int_0^\infty \int_0^x dy dF(x) - \int_{-\infty}^0 \int_x^0 dy dF(x). \end{aligned}$$

Now, since X is an absolutely integrable r.v. (i.e. $E[\mid X \mid] < \infty$,) it is possible to use Fubini's Theorem and interchange the order of integration. In this way

[2]This result has applications in survival analysis, for example. It can be used to compute the mean residual life function:

$$r(t) = E[T - t \mid T \geq t] = (1/S(t)) \int_t^\infty S(x)dx$$

and $S(t) = Pr[T > t]$ is the survival rate.

we find that

$$\begin{aligned}E[X] &= \int_0^\infty (\int_y^\infty dF(x))dy - \int_{-\infty}^0 (\int_{-\infty}^y dF(x))dy \\ &= \int_0^\infty \Pr[X \geq y]dy - \int_{-\infty}^0 \Pr[X \leq y]dy \\ &= \int_0^\infty (1 - F(y))dy - \int_{-\infty}^0 F(y)dy\end{aligned}$$

which provides a different and quicker proof of the first statement of the problem.

PROBLEM 1.1.2. Let X be an $\mathbf{R}^+$-valued random variable and let ψ be a monotonic, left continuous real valued random variable. Prove that

$$E[\psi(X)] = \psi(0) + \int_0^\infty (1 - F(y))\, d\psi(y)$$

where F is the cumulative distribution function for X.

SOLUTION. By definition, the expected value of $\psi(X)$ is given by

$$E[\psi(X)] = \int_0^\infty \psi(x)\, dF(x).$$

This integral is well defined since, by assumption, $F(x)$ is right-continuous while $\psi(x)$ is left-continuous. The last expression can also be written as

$$\begin{aligned}E[\psi(X)] &= \int_0^\infty \int_0^x d\psi(y)\, dF(x) + \int_0^\infty \psi(0)\, dF(x) \\ &= \psi(0) + \int_0^\infty \int_0^x d\psi(y)\, dF(x).\end{aligned}$$

Now, since

$$\int_0^\infty \int_0^x d\psi(y)\, dF(x) = \int_{[0,\infty)\times[0,\infty)} I_{\{(x,y):y\leq x\}}\, d\psi(y)\, dF(x)$$

and both F and ψ are monotonic, the integral above exists (finite or infinite) and, therefore, it is possible to use Fubini's Theorem. This gives:

$$\begin{aligned}E[\psi(X)] &= \psi(0) + \int_0^\infty \int_y^\infty dF(x)\, d\psi(y) \\ &= \psi(0) + \int_0^\infty (1 - F(y))\, d\psi(y)\end{aligned}$$

as we were supposed to prove.

Remark A. This problem contains the first result of Problem 1.1.1 as a special case. In fact, if $\psi(x) = x$ we have

$$E[X] = \int_0^\infty (1 - F(x))\, dx.$$

Remark B. The result just established can be used to show that if Z is any $\mathbf{R}^+$-valued random variable and s is any positive constant then

$$E[Z^s] = \int_0^\infty z^s\, dF(z) = s\int_0^\infty z^{s-1}(1 - F(z))\, dz.$$

PROBLEM 1.1.3. Let X be a $\mathbf{R}_+$-valued random variable defined on some probability space $(\Omega, \mathcal{F}, P)$. Suppose that for every $0 < q < 1$ there is $T = T(q)$ such that

$$P[X > 2t] \leq qP[X > t] \;\; \forall t > T.$$

Prove that all moments of X are finite.
Let Y be also a $\mathbf{R}_+$-valued random variable on the same probability space such that

$$P[Y > 2t] \leq (P[Y > t])^2 \;\; \forall t > 0.$$

Prove that there exists a number θ such that $E[e^{\lambda Y}] < \infty$ for $\lambda \in [0, \theta)$.

SOLUTION. Let p be any positive integer number and $T^0(p)$ be such that the inequality above holds for $q < 2^{-p}$. Then, using Remark B of Problem 1.1.2, for every $T > T^0(p)$ we can write:

$$\begin{aligned}
E[X^p] &= p\int_0^\infty x^{p-1}(1 - F(x))\, dx = p\int_0^\infty x^{p-1} P[X > x]\, dx \\
&= p\left[\int_0^T x^{p-1} P[X > x]\, dx + \sum_{i=1}^\infty \int_{2^{i-1}T}^{2^i T} x^{p-1} P[X > x]\, dx\right] \\
&\leq p\left[\int_0^T x^p P[X > x]\, dx + \sum_{i=1}^\infty (2^i T)^{p-1} P[X > 2^{i-1}T] \int_{2^{i-1}T}^{2^i T} dx\right] \\
&\leq p\left[\int_0^T x^p P[X > x]\, dx + \left(\frac{T}{2}\right)^p P[X > T]\left(\sum_{i=1}^\infty (2^p q)^{i-1}\right)\right] < \infty.
\end{aligned}$$

In fact, $\int_0^T x^p P[X > x]\, dx < \infty$, $(T/2)^p P[X > T] < \infty$, and $\sum_{i=1}^\infty (2^p q)^{i-1}$ is also finite since we selected T so that $q < 2^{-p}$.
Let $s \neq 0$ and let q be the number such that $P[Y > s] < q < 1$. Then we can prove by mathematical induction that $P[Y > 2^n s] \leq q^{2n}$. In fact, the statement is easily seen to hold for $n = 1$ since, by assumption, it is

$$P[Y > 2s] \leq (P[Y > s])^2 = q^2.$$

If we assume that the statement also holds for $n = m > 1$, we find that

$$P[Y > 2^{m+1}s] = P[Y > 2(2^m s)] \le (P[Y > 2^m s])^2$$
$$\le P[Y > 2s] \cdot P[Y > 2^m s] < q^2 \cdot q^{2m} = q^{2(m+1)}$$

which completes the induction. Now, for any $t > 0$ it is possible to find s such that $2^n s \le t < 2^{n+1}s$. Then

$$P[Y > t] \le P[Y > 2^n s] \le q^{2n}.$$

On the other hand, from the inequality $t < 2^{n+1}s$ it follows that $2n < t/(2s)$ and also $P[Y > t] \le q^{t/(2s)}$. It is now easy to check that if $\theta = -\frac{\log q}{2s}$ then $P[Y > t] \le e^{-\theta t}$. This proves the second statement of the problem as, in fact,

$$E[e^{\lambda Y}] = \int_0^\infty e^{\lambda t} P[Y > t]\, dt \le \int_0^\infty e^{(\lambda-\theta)t}\, dt < \infty.$$

Remark. The second part of this problem provides a condition for the existence of moment generating functions.

PROBLEM 1.1.4. Let X be a real-valued random variable with distribution function $F(x)$ and such that $E[X^2]$ exists. Prove that

$$E[X^2] = 2\int_0^\infty x[1 - F(x) + F(-x)]\, dx$$

and

$$\lim_{x\to\infty} x^2[1 - F(x) + F(-x)] = 0.$$

SOLUTION. According to the definition,

$$E[X^2] = \int_{-\infty}^\infty x^2\, dF(x) = \int_0^\infty x^2\, dF(x) + \int_{-\infty}^0 x^2\, dF(x)$$
$$= 2\left[\int_0^\infty \left(\int_0^x y\, dy\right) dF(x) - \int_{-\infty}^0 \left(\int_x^0 y\, dy\right) dF(y)\right]$$
$$= 2\left[\int_0^\infty \int_y^\infty dF(x) y\, dy - \int_{-\infty}^0 \int_{-\infty}^y dF(x) y\, dy\right]$$

where the order of integration has been changed according to Fubini's Theorem. This is clearly possible as both integrals on the right-hand side exist either finite or infinite. Then, with a suitably chosen change of variable, we have also

$$E[X^2] = 2\left[\int_0^\infty y[1 - F(y)]\, dy - \int_\infty^0 y[F(y)]\, dy\right]$$
$$= 2\left[\int_0^\infty y[1 - F(y)]\, dy + \int_\infty^0 -z[F(-z)]\, dz\right]$$
$$= 2\left[\int_0^\infty y[1 - F(y) + F(-y)]\, dy\right]$$

and this proves the first statement of the problem. Next, according to the formula just found, one has

$$E[X^2] = \lim_{x\to\infty} \int_0^x 2y[1 - F(y) + F(-y)]\, dy$$

which, using a simple integration by parts, is equivalent to

$$\begin{aligned}
E[X^2] &= \lim_{x\to\infty} x^2[1 - F(x) + F(-x)] - \int_0^\infty y^2\, d[1 - F(y) + F(-y)] \\
&= \lim_{x\to\infty} x^2[1 - F(x) + F(-x)] + \int_0^\infty y^2\, dF(y) - \int_0^\infty y^2\, dF(-y) \\
&= \lim_{x\to\infty} x^2[1 - F(x) + F(-x)] + \int_0^\infty y^2\, dF(y) + \int_{-\infty}^0 w^2\, dF(w) \\
&= \lim_{x\to\infty} x^2[1 - F(x) + F(-x)] + E[X^2]
\end{aligned}$$

and the second statement of the problem follows easily.

PROBLEM 1.1.5. Prove that for every distribution function, F and for every $a > 0$

$$\int_{\mathbf{R}} (F(x + a) - F(x))dx = a.$$

SOLUTION. The integral exists (finite or infinite) as, by assumption, we have $a > 0$ and therefore $F(x + a) - F(x) \geq 0$. Letting Q be the probability measure associated with F, the original integral can be rewritten as

$$\int_{\mathbf{R}} (F(x + a) - F(x))\, dx = \int_{\mathbf{R}} \int_x^{x+a} Q(dy)dx.$$

Using Fubini's Theorem and interchanging the order of integration we find

$$\begin{aligned}
\int_{\mathbf{R}} \int_x^{x+a} Q(dy)dx &= \int_{\mathbf{R}} \int_{y-a}^{y} dxQ(dy) \\
&= \int_{\mathbf{R}} a\, Q(dy) = aQ(\mathbf{R}) = a \cdot 1 = a.
\end{aligned}$$

PROBLEM 1.1.6. Prove that $E[\mid X \mid^r] < \infty$ iff $\int z^{r-1} P[\mid X \mid \geq z]dz < \infty$.

SOLUTION. Let $Y = \mid X \mid$ and let F be the distribution function for Y. Then, $E[\mid X \mid^r] = E[Y^r]$ and thus, using the definition of expectation, we have that

$$E[Y^r] = \int_0^\infty y^r\, dF(y) = \int_0^\infty \int_0^y r z^{r-1}\, dz dF(y).$$

Now, if we assume that $E[Y^r] < \infty$, the integral above also exists and, therefore, we can use Fubini's Theorem to interchange the order of integration. In this way we find that

$$\begin{aligned}\infty > E[Y^r] &= \int_0^\infty (r \int_z^\infty dF(y)) z^{r-1}\, dz \\ &= r \int_0^\infty z^{r-1} P(Y \geq z)\, dz.\end{aligned}$$

Since $Y = \mid X \mid$ and r is finite, this proves the if part. The only if part is trivial.

1.2 Inequalities

PROBLEM 1.2.1. [3] Let X have distribution function F. Then

(a) $E[X^+] < \infty$ if and only if $\int_\alpha^\infty (-\log F(t))dt < \infty$ for some α;

(b) if α and $F(\alpha)$ are positive,

$$\int_{X>\alpha} X dP \leq \alpha(-\log F(\alpha)) + \int_\alpha^\infty (-\log F(t))dt \leq \frac{1}{F(\alpha)} \int_{X>\alpha} X dP.$$

SOLUTION. To prove statements (a) and (b) we need first to state and prove a couple of lemmas.

Lemma 1. $1 - x \leq -\log x, \forall x \in (0, 1]$.

Proof. Let $H(x) = \log(x) - x + 1$. Then, $\lim_{x \to 0+} H(x) = -\infty$, $H(1) = 0$ and $H'(x) = -1 + 1/x$ is strictly positive for every choice of $x \in (0, 1)$. This proves that $H(x) \leq 0$ in $(0, 1]$ and completes the proof of the lemma. □

Lemma 2. $-\log x \leq (1 - x)/F(\alpha), \forall x \in [F(\alpha), 1]$.

Proof. Let $K(x) = -\log x + (x-1)/F(\alpha)$. Since $K(1) = 0$ and $K'(x) = 1/F(\alpha) - 1/x > 0$ for any $x \in (F(\alpha), 1]$, it follows that $K(x) \leq 0$ $\forall x \in [F(\alpha), 1]$ and therefore the result stated in the lemma holds. □

Since $F(t)$ is a number between 0 and 1, using Lemma 1 we find that $1 - F(t) \leq -\log F(t)$ and therefore

$$E[X^+] = \int_0^\infty [1 - F(t)]\, dt \leq \alpha + \int_\alpha^\infty [1 - F(t)]\, dt \leq \alpha + \int_0^\infty [-\log F(t)]\, dt$$

[3] This is Problem 21.12 in P. Billingsley, Probability and Measure, 1991; p. 288.

thus proving that $\int_\alpha^\infty(-\log F(t))dt < \infty$ implies $E[X^+] < \infty$ as α is finite. Replacing x by $F(t)$ in Lemma 2 we get that $-\log F(t) \le (1-F(t))/F(\alpha)$, $\forall t \in [\alpha, 1]$ using the fact that $F(\cdot)$ being a distribution function is not decreasing[4]. The last inequality yields

$$\int_\alpha^\infty (-\log F(t))dt \le \int_\alpha^\infty \frac{1-F(t)}{F(\alpha)}dt$$

which implies

$$F(\alpha)\int_\alpha^\infty (-\log F(t))dt \le \int_\alpha^\infty (1-F(t))dt \le \int_0^\infty (1-F(t))dt = E[X^+].$$

Hence $E[X^+] < \infty$ implies $\int_\alpha^\infty(-\log F(t))dt < \infty$ as $F(\alpha) \in [0,1]$ and proves that the first statement of the problem is correct.

As for the second statement, using the result in problem 1 and replacing X with $X \cdot I_{[X>\alpha]}$, we have

$$\int_{X>\alpha} XdP = \alpha P[X>\alpha] + \int_\alpha^\infty P[X>t]dt.$$

Since $P[X>t] = 1-F(t) \le -(\log F(t))$, the equality above can be transformed into the following:

$$\int_{X>\alpha} XdP \le \alpha(-\log F(\alpha)) + \int_\alpha^\infty (-\log F(t))dt$$

which proves the first of the two inequalities. Using Lemma 2, we find also

$$\begin{aligned}\alpha(-\log F(\alpha)) + \int_\alpha^\infty (-\log F(t))dt = \int_\alpha^\infty (-\log F(t))dt \le \\ \le \frac{1}{F(\alpha)}\int_\alpha^\infty [1-F(t)]dt = \frac{1}{F(\alpha)}\int_{X>\alpha} XdP\end{aligned}$$

which is what we were trying to prove.

PROBLEM 1.2.2. Let $f, g : \mathbf{R} \mapsto \mathbf{R}$ be either both non-increasing or both non-decreasing. Let also X and Y be $\mathbf{R}$-valued random variables such that $\mathcal{L}(X,Y) = \mathcal{L}(Y,X)$ (X and Y are exchangeable). Show that:

(a)
$$E[f(X)g(X)] \ge E[f(X)g(Y)]$$

(b) In the above set-up, prove that

$$E[f(X)g(X)] \ge E[f(X)]E[g(X)]$$

[4] Actually, unless the distribution function is continuous, the inequality above might be true for all $t \in [\beta, 1]$ with $\beta < \alpha$.

provided that the expectations are defined.

SOLUTION. To prove the first inequality, we start by assuming that all expectations are meaningful. Then we use the fact that

$$(f(X) - f(Y))(g(X) - g(Y)) \geq 0$$

as f and g are both non-increasing or both non-decreasing. This implies that

$$\int_{\mathbf{R}^2} (f(x) - f(y))(g(x) - g(y))\, Q(dx, dy) \geq 0.$$

The inequality above can also be written as

$$\int_{\mathbf{R}^2} f(x)g(x)\, Q(dx, dy) + \int_{\mathbf{R}^2} f(y)g(y)\, Q(dx, dy)$$
$$- \int_{\mathbf{R}^2} f(y)g(x) Q(dx, dy) - \int_{\mathbf{R}^2} f(x)g(y)\, Q(dx, dy) \geq 0$$

or

$$\int_{\mathbf{R}} f(x)g(x)\, Q_X(dx) + \int_{\mathbf{R}} f(y)g(y)\, Q_Y(dy)$$
$$- \int_{\mathbf{R}^2} f(y)g(x)\, Q(dx, dy) - \int_{\mathbf{R}^2} f(x)g(y)\, Q(dx, dy) \geq 0$$

where $Q_X(dx) = \int_{\mathbf{R}} Q(dx, dy)$ and $Q_Y(dy)$ is defined similarly.
At this point we can use the assumption of exchangeability. The fact that $\mathcal{L}(X, Y) = \mathcal{L}(Y, X)$ implies that

$$\int_{\mathbf{R}} \int_{\mathbf{R}} f(x)g(y)\, Q(dx, dy) = \int_{\mathbf{R}} \int_{\mathbf{R}} f(y)g(x)\, Q(dy, dx)$$

and, since $\mathcal{L}(X, Y) = \mathcal{L}(Y, X)$ implies[5] $\mathcal{L}(X) = \mathcal{L}(Y)$, we find also

$$\int_{\mathbf{R}} f(x)g(x)\, Q_X(dx) = \int_{\mathbf{R}} f(y)g(y)\, Q_Y(dy).$$

Using these two facts, we can finally re-write the original inequality as

$$2 \int_{\mathbf{R}} f(x)g(x)\, Q_X(dx) \geq 2 \int_{\mathbf{R}} \int_{\mathbf{R}} f(x)g(y)\, Q(dx, dy)$$

[5]This follows from the following argument. Assume that X and Y are exchangeable r.v.'s, this means that $Q(dx, dy) = Q(dy, dx)$. Then, we have

$$Q_Y(dy) = \int_{\mathcal{X}} Q(dx, dy) = \int_{\mathcal{Y}} Q(dy, dx) = Q_X(dx)$$

where $\mathcal{X}$ and $\mathcal{Y}$ are the domains for X and Y, respectively. It is clear that the two domains are the same.

or

$$E[f(X)g(X)] \geq E[f(X)g(Y)]$$

as we were supposed to prove.

To prove the second inequality we need to assume again that the expectations are defined so that the inequality is meaningful. Then, it is possible to prove the new inequality using the first. In fact, let T and S be two indepedent r.v.'s both with the same distribution as X. Being independent, these two r.v.'s are also exchangeable and, therefore, the first inequality applies. Thus,

$$E[g(T)f(T)] \geq E[f(T)g(S)]$$

which, using the fact that T and S are independent and both have the same distribution as X, can also be written as

$$E[f(T)g(T)] \geq E[f(T)]E[g(S)]$$

or

$$E[f(X)g(X)] \geq E[f(X)]E[g(X)].$$

PROBLEM 1.2.3. Let X be a continuous random variable with density function $f(x)$. Let g and g' be real functions on $\mathbf{R}$ such that g is the undefined integral of g' and assume that $Var[g(X)] < \infty$. Prove that

$$Var[g(x)] \leq \int_0^\infty \int_t^\infty xf(x)[g'(t)]^2\,dxdt - \int_0^\infty \int_{-\infty}^t xf(x)[g'(t)]^2\,dxdt.$$

Then, assuming that $X \sim N(0,1)$ and $g(x)$ satisfies the above conditions, prove that[6]

$$Var[g(x)] \leq E[g'(x)]^2$$

with equality iff $g(x)$ is linear.

SOLUTION. Since $g(x) = \int_0^x g'(t)\,dt + g(0)$, it is easily seen that $g(x)$ and $\int_0^x g'(t)dt$ differ only by a constant. This fact allows us to write

$$\mathrm{Var}[g(x)] = \mathrm{Var}\left[\int_0^x g'(t)\,dt\right] \leq E\left[\int_0^x g'(t)\,dt\right]^2.$$

Applying the Cauchy-Schwarz Inequality to the integral inside the square brackets gives:

$$\left[\int_0^x 1\cdot g'(t)\,dt\right]^2 \leq \int_0^x 1^2dt \int_0^x [g'(t)]^2\,dt$$

[6]This is part of the more general inequality (Chernoff's Inequality)

$$E^2[g'(x)] \leq Var[g(x)] \leq E[g'(x)]^2.$$

and, thus, we have

$$\begin{aligned}\operatorname{Var}[g(x)] &\le E\left[\int_0^x 1^2\,dt \int_0^x [g'(t)]^2\,dt\right] = E\left[x\cdot\int_0^x [g'(t)]^2\,dt\right]\\ &= \int_{-\infty}^{\infty} xf(x)\int_0^x [g'(t)]^2\,dtdx = \int_0^{\infty}\int_0^x xf(x)[g'(t)]^2\,dtdx\\ &- \int_{-\infty}^0\int_0^x xf(x)[g'(t)]^2\,dtdx.\end{aligned}$$

As by assumption the expected values involved exist and are finite, it is possible to use Fubini's Theorem to change the order of integration. This gives

$$\operatorname{Var}[g(x)] \le \int_0^{\infty}\int_t^{\infty} xf(x)[g'(t)]^2\,dxdt - \int_{-\infty}^0\int_{-\infty}^t xf(x)[g'(t)]^2\,dxdt$$

which proves the first inequality. The second inequality follows from the one just proved. In fact, it suffices to take $f(x) = \phi(x) \equiv 1/(2\pi)^{1/2}e^{-x^2/2}$ and use the (easily proved) facts

$$\int_t^{\infty} x\phi(x)dx = \phi(t), \quad \int_{-\infty}^t x\phi(x)dx = -\phi(t).$$

It is then possible to write for this special case that

$$\operatorname{Var}[g(x)] \le \int_0^{\infty} [g'(t)]^2\phi(t)\,dt + \int_{-\infty}^0 [g'(t)]^2\,dt = E[(g'(t))^2].$$

The last assertion can be proved using an expansion of g(x) in terms of orthonormalized Hermite polynomials and the details are available in Chernoff's paper *A note on an inequality involving the normal distribution*, Ann. Probab., (1981) **9** 533-35. A generalization of this inequality can be found in Cacoullos's paper, *On upper and lower bounds for the variance of a function of a random variable.*, Ann. Probab., (1982) **10** 799-809.

1.3 Quantiles, Median, Mode and Inequalities

PROBLEM 1.3.1. Let F be a distribution function on $\mathbf{R}^1$ such that $\int_{\mathbf{R}} \mid x \mid F(dx) < \infty$.

(a) For each $b \in \mathbf{R}^1$, show that

$$\int_{\mathbf{R}} \mid x - b \mid F(dx) = \int_{-\infty}^b F(x)dx + \int_b^{\infty}(1 - F(x))dx.$$

(b) Let $m_1 = \inf\{x : F(x) > 0.5\}$ and let $m_2 = \sup\{x : F(x) < 0.5\}$. Any number in $[m_2, m_1]$ is a median of F. Let $m \in [m_2, m_1]$. Show that $F(m) \geq 0.5$, $F(m^-) \leq 0.5$ and if F is continuous $F(m) = 0.5$.

(c) Show that

$$\int_{\mathbf{R}} | x - b | F(dx) \geq \int_{\mathbf{R}} | x - m | F(dx)$$

for all $b \in \mathbf{R}$.

(d) Let $X_1, X_2, \ldots, X_n$ be iid r.v.'s from a continuous distribution function G and let G_n be the empirical distribution function. Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be the order statistics. With $F = G_n$ in the previous parts, show that one possibility for m is

$$m = \begin{cases} X_{((n+1)/2)} & \text{if } n \text{ is odd,} \\ \frac{X_{(n/2)} + X_{(n/2+1)}}{2} & \text{if } n \text{ is even.} \end{cases}$$

SOLUTION. Since by assumption $\int_{-\infty}^{\infty} | x | F(dx) < \infty$, the integral exists, a simple application of the Fubini-Tonelli Theorem yields

$$\int_{-\infty}^{\infty} | x - b | F(dx) = \int_{\mathbf{R}} (\int_{\mathbf{R}} I_{[x \leq t \leq b, x \leq b]} dt) F(dx) + \int_{\mathbf{R}} (\int_{\mathbf{R}} I_{[b \leq t \leq x, x \geq b]} dt) F(dx)$$
$$= \int_{-\infty}^{b} F(t) dt + \int_{b}^{\infty} (1 - F(t)) dt.$$

Of course, the same result could be established using integration by parts, but this would require a lot of algebra and several $\infty \cdot 0$ forms are involved.

Let $A = \{x : F(x) < 0.5\}$. Then $A^C = \{x : F(x) \geq 0.5\}$ is a closed set since F is a right continuous function. Then $m_2 \in A^C$ and $F(m_2) \geq 0.5$, so that $F(m) \geq F(m_2) \geq 0.5$. Similarly, if we let $B = \{x : F(x) > 0.5\}$, we find that $m_1 \in \bar{B}$, the closure of B. If B is either left open or left closed it follows that $m_1^- \in B^C$ so $F(m_1^-) \leq 0.5$.

Let's start with the case $b > m$. In this case the results from (a) and (b) imply

$$\int | x - b | F(dx) - \int | x - m | F(dx) = \int_{-\infty}^{b} F(x) dx + \int_{-\infty}^{b} (1 - F(x)) dx$$
$$- (\int_{-\infty}^{m} F(x) dx + \int_{-\infty}^{m} (1 - F(x)) dx = \int_{m}^{b} F(x) dx - \int_{m}^{b} (1 - F(x)) dx$$
$$= \int_{m}^{b} (2F(x) - 1) dx \geq 0.$$

The other case is similar.

If G_n is the empirical distribution of $X_1, X_2, \ldots, X_n$, a random sample from a continuous distribution function G, it is easily checked that

$$G_n(t) = \frac{1}{n}\sum_{i=1}^{n} I_{(0,\infty)}(X_i - t).$$

Now, if $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ are the order statistics, it follows

$$G_n(t) = \begin{cases} 0 & \text{if } x < X_{(1)}; \\ 1/n & \text{if } X_{(1)} \le x < X_{(2)}, \\ 2/n & \text{if } X_{(2)} \le x < X_{(3)}, \\ \cdots & \ldots, \\ (n-1)/n & \text{if } X_{(n-1)} \le x < X_{(n)}, \\ 1 & \text{if } x \ge X-(n). \end{cases}$$

Assuming that G is a continuous distribution function means that the probability that two or more of the X_i's are the same is zero and therefore the definition of $G_n(x)$ given above is correct. Then for n odd, $G_n(X_{((n+1)/2)}) = (n+1)/(2n)$ and $G_n(X^-_{((n+1)/2)}) = (n-1)/(2n)$. Hence $m_1 = m_2 = X_{((n+1)/2)}$.. When n is even, $G_n(X_{(n/2)}) = 0.5$ and $G_n(X^-_{(n/2)}) = 0.5 - 1/n$. Hence $m_2 = X_{(n/2)}$. Also note that $G_n(X_{(n/2+1)} = 0.5 + 1/n$ and $G_n(X^-_{(n/2+1)}) = 0.5$ so that $m_1 = X_{(n/2+1)}$. Since any number in $[m_2, m_1]$ is a median of F, then

$$m = \begin{cases} X_{((n+1)/2)} & \text{if } n \text{ is odd}, \\ \frac{X_{(n/2)}+X_{(n/2+1)}}{2} & \text{if } n \text{ is even}. \end{cases}$$

is clearly one possibility.

PROBLEM 1.3.2. Let X be a non-negative unimodal positively skewed random variable with a continuous pdf, $f(x)$ and cdf $F(x)$. Assume that $M > 0$ is the mode of this random variable. Denote the median and the mean of X by m and μ, respectively. Then, prove[7] that $M < m < \mu$.

SOLUTION. Let

$$g(x) = \begin{cases} f(x) & \text{if } 0 \le x \le M \\ f(2M - x) & \text{if } M < x \le 2M \end{cases}$$

so that $g(x)$ is the rotation of the pdf $f(x)$ from 0 to M around $x = M$. If $g(x) \le f(x)$ for $M < x \le 2M$, with the strict inequality holding for at least one x, then it must be $m > M$. In fact, assume that $m \le M$. Then it must be

$$1 = \int_0^\infty f(x)dx > \int_0^{2M} g(x)dx > 2\int_0^m f(x)dx = 1.$$

[7]This proof is due to R.A. Groeneveld and G. Meeden, The American Statistician, 1977, Vol. 31, N.3.

This is a contradiction and hence $m > M$.

Now, define the function h as follows:

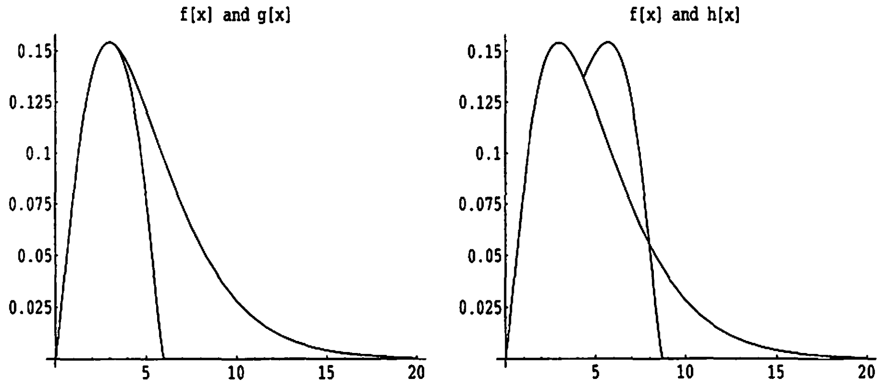

Figure 1.1: $g(x)$ and $h(x)$ in the case of a positively skewed random variable.

$$h(x) = \begin{cases} f(x) & \text{if } 0 \le x \le m, \\ f(2m - x) & \text{if } m < x \le 2m. \end{cases}$$

The function $h(x)$ defines a rotation of $f(x)$ from 0 to m around m. It is easily checked that it is a density function for some random variable, say Y, symmetric around m and such that $E[Y] = m$. Let $H(x)$ be its cdf. Then $H(x) = F(x)$ for $0 \le x \le m$. If $f(x)$ is a density function for which there exists a unique $x_1 > m$ such that $h(x) > f(x)$ for $m < x < x_1$, $h(x_1) = f(x_1)$, and $h(x) < f(x)$ for $x > x_1$, then

$$1 - F(x) \ge 1 - H(x)$$

for all x, with a strict inequality when $x > m$. For $x \le m$ the inequality is obvious. When $m < x \le x_1$ it follows that $F(x) < H(x)$. Hence, for the positive random variables X and Y we have

$$\mu = \int_0^\infty (1 - F(x))dx > \int_0^\infty (1 - H(y))dy = m.$$

Therefore, we have established that $M < m < \mu$.

Remark. To apply this result, the inequalities between $f(x), g(x)$ and $f(x), h(x)$ are easier to verify when the transformation $x = y + M$ is used. In this case to show that e.g. $g(x) \le f(x)$ for $M < x \le 2M$ it suffices to show that

$$\frac{g(x)}{f(x)} = \frac{f(M - y)}{f(M + y)} = r(y) \le 1$$

for $0 \le y \le M$, with strict inequality holding for at least one y.

1.4 Moments and Indicator Functions

PROBLEM 1.4.1. Let $X_1, X_2, \ldots, X_n$ be iid random vectors in $\mathbf{R}^k$ and let B be any Borel set in $\mathbf{R}^k$. Let $p(B) = P[X_1 \in B]$. Define random variables $Y_i(B) = I_{\{X_i \in B\}}$ for $i = 1, 2, \ldots, n$ and let $Z_n(B) = (1/n)\sum_{i=1}^n Y_i(B)$. Note if $k = 1$ and $B = (-\infty, t]$, then $Z_n(B) = \hat{F}_n(t)$, the empirical distribution function.

(a) Fix B and let $Z_n = Z_n(B)$ and $p = p(B)$. Show that $E[Z_n] = p$ and $Var[Z_n] = n^{-1}p(1-p)$.

(b) For two Borel sets B_1 and B_2, find $Cov[Z_n(B_1), Z_n(B_2)]$.

SOLUTION. As the X_i's are identically distributed, $p = P[X_i \in B]\, i = 1, 2, \ldots, n$. Using the assumptions of the problem, it is easily seen that when $k = 1$ and $B = (-\infty, t]$,

$$Z_n((-\infty, t]) = \frac{1}{n}\sum_{i=1}^n Y_i(B) = \frac{\# X_i^{\prime} s \in (-\infty, t]}{n} \stackrel{def}{=} \hat{F}_n.$$

In addition, it is also

$$E[Z_n(B)] = \frac{1}{n}\sum_{i=1}^n E[Y_i(B)] \quad \text{as} \quad Y_i(B) = \begin{cases} 1 & \text{if } X_i \in B, \\ 0 & \text{otherwise.} \end{cases}$$

and thus it must be

$$E[Z_n(B)] = \frac{1}{n}\sum_{i=1}^n [1 \cdot P[X_i \in B] + 0 \cdot P[X_i \in B^C]] = \frac{np(B)}{n} = p(B) = p.$$

In order to compute the variance of $Z_n(B)$ it is necessary to compute its second moment first, i.e.:

$$E[Z_n^2(B)] = E[((1/n)\sum_{i=1}^n Y_i(B))^2] = E[(1/n^2)\sum_{i=1}^n Y_i^2(B) + \sum_{i=1}^{j}\sum_{j=1}^{n} Y_i(B)Y_j(B)].$$

Since in this case we have

$$Y_i^2 = \begin{cases} 1 & \text{if } X_i \in B, \\ 0 & \text{otherwise.} \end{cases} \quad Y_iY_j = \begin{cases} 1 & \text{if } X_i \in B \text{ and } X_j \in B, \\ 0 & \text{otherwise.} \end{cases}$$

it follows that $Y_i^2 = 1$ with probability p and as we are working with a random sample, it is also $Y_iY_j = 1$ with probability p^2. These facts yield

$$E[Z_n^2(B)] = \frac{1}{n^2}(np + n(n-1)p^2) = \frac{p}{n} + \frac{n-1}{n}p^2$$

and hence

$$Var[Z_n(B)] = \frac{p}{n} + \frac{n-1}{n}p^2 - p^2 = \frac{p(1-p)}{n}.$$

Let now $p_1 = P[X_1 \in B_1]$, $p_2 = P[X_1 \in B_2]$ where B_1 and B_2 are two Borel subsets of $\mathbf{R}^k$ and let also $p_{12} = P[X_1 \in B_1 \cap B_2]$. Using again the fact that we have a random sample, we have that $p_{12} = P[X_i \in B_1 \cap B_2]$, $i = 1, 2, \ldots, n$. Then

$$Cov[Z_n(B_1), Z_n(B_2)] = E[Z_n(B_1)Z_n(B_2)] - E[Z_n(B_1)]E[Z_n(B_2)]$$

where, using (a),

$$E[Z_n(B_1)] = p_1; \;\; E[Z_n(B_2)] = p_2.$$

On the other hand,

$$\begin{aligned} E[Z_n(B_1)Z_n(B_2)] &= E\left[\frac{1}{n}\sum_{i=1}^{n} Y_i(B)\right] \cdot \left[\frac{1}{n}\sum_{i=1}^{n} Y_i(B)\right] \\ &= E\left[\frac{1}{n^2}\left(\sum_{i=1}^{n} Y_i(B)Y_i(B_2) + \sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n} Y_i(B)Y_j(B_2)\right)\right]. \end{aligned}$$

Now, we have

$$Y_i(B_1)Y_i(B_2) = \begin{cases} 1 & \text{if } X_1 \in B_1 \cap B_2, \\ 0 & \text{otherwise.} \end{cases}$$

for all $i = 1, 2, \ldots$ and

$$Y_i(B_1)Y_j(B_2) = \begin{cases} 1 & \text{if } X_1 \in B_1 \text{ and } X_j \in B_2, \\ 0 & \text{otherwise} \end{cases}$$

for all $i, j = 1, 2, \ldots$ with $i \neq j$. The last two equalities together yield

$$E[Y_i(B_1)Y_i(B_2)] = p_{12}, \quad i = 1, 2, \ldots, n$$

and

$$E[Y_i(B_1)Y_j(B_2) = p_1 p_2, \quad i, j = 1, 2, \ldots, n; i \neq j.$$

Thus, we have found that

$$E[Z_n(B_1)Z_n(B_2)] = \left[\frac{1}{n^2}(np_{12} - n(n-1)p_1p_2\right] = \frac{p_{12} + (n-1)p_1p_2}{n}$$

and finally, combining the last two expressions,

$$Cov[Z_n(B_1), Z_n(B_2)] = \frac{p_{12}}{n} + \frac{(n-1)p_1p_2}{n} - p_1p_2 = \frac{p_{12} - p_1p_2}{n}.$$

PROBLEM 1.4.2. Consider a deck of n cards labeled 1 through n. Assume that the deck is well-shuffled (i.e., each of the n! arrangements of the deck is equally likely). Let N be the number of cards whose position in the shuffled deck (counting from the top) is at least as large as its label. Obtain formulas for the mean and variance of N.

SOLUTION. This problem provides an interesting application where indicator functions are a very practical tool in computing moments for relatively complicated random variables. In fact, let

$$I_i = \begin{cases} 1 & \text{if i-th card has label } j \leq i; \\ 0 & \text{otherwise.} \end{cases}$$

Then it is easily verified that the variable N can be written as

$$N = \sum_{i=1}^{n} I_i,$$

so that

$$E[N] = \sum_{i=1}^{n} P[I_i = 1].$$

Now, $P[I_i = 1] = i/n$ and, therefore, $E[N] = \sum_{i=1}^{n}(i/n) = (n+1)/2$. To find the variance of N one can use the identity

$$Var[N] = E[N^2] - (E[n])^2.$$

and the linearity of expectation to get

$$\begin{aligned} &= E\left[\sum_{i=1}^{n} I_i\right]^2 - \left(\frac{n+1}{2}\right)^2 \\ &= \sum_{i=1}^{n} E[I_i^2] + 2\sum_{i<j} E[I_i I_j] - \left(\frac{n+1}{2}\right)^2 \\ &= \left(\frac{n+1}{2}\right) + 2\sum_{i<j} E[I_i I_j] - \left(\frac{n+1}{2}\right)^2. \end{aligned}$$

Since

$$I_i I_j = \begin{cases} 1 & \text{if i-th card has label } h \leq i \text{ and j-th card has label } k \leq j;\ i < j \\ 0 & \text{otherwise} \end{cases}$$

it is easy to verify that

$$P[I_i \cdot I_j = 1] = \frac{i \cdot (j-1) \cdot (n-2)!}{n!} = \frac{i \cdot (j-1)}{n(n-1)}$$

and hence

$$Var[n] = \left(\frac{n+1}{2}\right) + 2\sum_{i<j} \frac{i(j-1)}{n(n-1)} - \left(\frac{n+1}{2}\right)^2.$$

1.5 Conditioning and Moments

PROBLEM 1.5.1. Suppose $X_1, X_2, \ldots,$ are i.i.d. random variables with distribution function F. Suppose $N \sim$ Poisson(λ) independent of the X's. Let $Y = \sum_{j=1}^{N} X_j$ where $Y = 0$ if $N = 0$.

(a) Show that $E[| Y |] < \infty$ iff $E[| X_1 |] < \infty$.

(b) Assume $E[| X_1 |] < \infty$. Find $E[Y]$ and the characteristic function of Y.

SOLUTION. We start by proving that $E[| Y |] < \infty \Rightarrow E \mid X_1 |] < \infty$. In fact, we know that if $E[| Y |] < \infty$, and then, since $E[| Y |] = E_N[E[| Y || N]]$, we have

$$\infty > E[| Y || N = 1]P[N = 1] = E[| X_1 |]P[N = 1].$$

This implies that $E[| X_1 |] < \infty$ as $P[N = 1] \in (0,1)$. since $N \sim$ Poisson(λ). To prove that $E[| X_1 |] < \infty \Rightarrow E[| Y |] < \infty$ we notice first that

$$[| Y |] \mid N = n = \mid \sum_{i=1}^{n} X_i \mid \leq \sum_{i=1}^{n} \mid X_i \mid; \quad n = 1, 2, 3, \ldots$$

and, using properties of expectation,

$$E[| Y || N = n] = E \mid \sum_{i=1}^{n} X_i \mid \leq \sum_{i=1}^{n} E \mid X_i \mid, \quad n = 1, 2, 3, \ldots$$

As by assumption the X_i's are i.i.d. r.v.'s, one can write $\sum_{i=1}^{n} \mid X_i \mid = n\, E\, [| \ X_1 \ |]$ and thus

$$E[| Y |] = \lim_{n\to\infty} \sum_{j=0}^{n} E[| Y || N = j]P[N = j] \leq \lim_{n\to\infty} \sum_{j=0}^{n} jE[| X_1 |]P[N = j]$$

$$= E[| X_1 |] \sum_{j=1}^{\infty} jP[N = j] = E[| X_1 |] \sum_{j=0}^{\infty} jP[N = j] = E[| X_1 |] \cdot E[N].$$

Therefore, since $N \sim$ Poisson(λ), $E[N] = \lambda > 0$ and finite, it follows that $E \mid Y \mid < \lambda E[| \ X_1 \ |] < \infty$ which implies $E[| Y |] < \infty$ and completes the proof of (a).

Assuming that $E[| X_1 |] < \infty$ we know from (a) that $E[| Y |] < \infty$ and, in addition, we know that $E[| Y |] < \infty$ implies that $E[Y]$ exists. Then, we can use the hierarchy of models to compute $E[Y]$; that is, using the fact that the X_i's are i.i.d. r.v.'s and they are independent of N,

$$E[Y] = E_N[E[Y \mid N]] = E_N[E[\sum_{i=1}^{N} X_i]] = E_N[N \cdot E[X_1]] = \lambda E[X_1].$$

Without the explicit knowledge of the distribution function F there is not very much one can do to compute $E[Y]$. Nonetheless, the characteristic function of Y is given by

$$\psi_Y(t) = E_N[E[e^{itY} \mid N]] = E_N[E[e^{it(\sum_1^N X_i)}]] = E_N[\psi_{\sum_1^n X_i}(t)]$$

and, as the X_i's are i.i.d.,

$$= E_N[(\psi_{X_1}(t))^n] = E_N[e^{N \log \psi_{X_1}(t)}] = M_N[\log[\psi_{X_1}(t)]]$$

where $M_N(\cdot)$ represents the moment generating function of the r.v. N and $\psi_{X_1}(\cdot)$ the characteristic function of the r.v. X_1. Now, since $N \sim$ Poisson(λ), we have

$$\psi_Y(t) = M_N[\log \psi_{X_1}(t)] = \exp\{\lambda[e^{\log \psi_{X_1}(t)} - 1]\} = \exp\{\lambda(\psi_{X_1}(t) - 1)\}.$$

PROBLEM 1.5.2. Let $X_1, X_2, \ldots$ be a sequence of independent random variables having the same distribution exponential(λ) and mean $1/\lambda$. Let $S_n = X_1 + X_2 + \ldots + X_n$, $n = 1, 2, \ldots$. Compute

$$E\left[\frac{S_1 + S_2 + \ldots + S_n}{n} \mid S_{n+1} = a\right].$$

SOLUTION. First we write the required conditional expectation in a form that is easier to deal with. To this purpose, using the linearity property of conditional expectation, we have

$$\begin{aligned} E\left[\frac{S_1 + S_2 + \ldots + S_n}{n} \mid S_{n+1} = a\right] &= \frac{1}{n}\sum_{k=1}^{n} E[S_k \mid S_{n+1} = a] \\ &= \frac{1}{n}\sum_{k=1}^{n}(n - k + 1)E[X_k \mid S_{n+1} = a]. \end{aligned}$$

Now,

$$X_i \sim \text{exponential}(\lambda) \Rightarrow f_{X_i}(x) = \lambda e^{-\lambda x} \cdot I_{[0,\infty)}(x)$$

and

$$S_n \sim \text{gamma}(n, \lambda) \Rightarrow f_{S_n}(y) = \frac{\lambda^n}{\Gamma(n)} y^{n-1} e^{-\lambda y} \cdot I_{[0,\infty)}(y); \quad n = 1, 2, \ldots$$

It is easily seen that $f_{X_i|S_{n+1}=a}(x)$ is the same for all values of $i = 1, 2, \ldots, n$ given that the X_i's are i.i.d. random variables and, thus, it suffices to compute $E[X_1 \mid S_{n+1} = a]$. Now,

$$\begin{aligned} f_{X_1,X_2,\ldots,X_n|S_{n+1}=a}(x_1, x_2, \ldots, x_n) &= \frac{f_{X_1,X_2,\ldots,X_n,S_{n+1}=a}(x_1, x_2, \ldots, x_n, a)}{f_{S_{n+1}}(a)} \\ &= \frac{n!}{a^n} \cdot I_{\{0 \le x_1 + x_2 + \ldots + x_n \le a\}}(x_1, x_2, \ldots, x_n). \end{aligned}$$

Integrating out $x_2, x_3, \ldots, x_n$ yields

$$\begin{aligned} & f_{X_1|S_{n+1}=a}(x_1) \\ &= \int_0^{a-x_1} dx_2 \ldots \int_0^{a-x_1-x_2-\ldots-x_{n-1}} f_{X_1,X_2,\ldots,X_n|S_{n+1}=a}(x_1, x_2, \ldots, x_n)\, dx_n \\ &= \frac{n}{a}\left(1 - \frac{x_1}{a}\right)^{n-1} \cdot I_{\{0 \le x_1 \le a\}}(x_1). \end{aligned}$$

The expected value of $E[X_1 \mid S_{n+1} = a]$ is then given by

$$E[X_1 \mid S_{n+1} = a] = n \int_0^a \frac{x}{a}\left(1 - \frac{x}{a}\right)^{n-1} dx = \frac{a}{n+1}.$$

Thus,

$$E\left[\frac{S_1 + S_2 + \ldots + S_n}{n} \mid S_{n+1} = a\right] = \frac{a}{2}.$$

PROBLEM 1.5.3. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. r.v.'s such that $E[|X_1|] < \infty$.. Let $S_n = X_1 + X_2 + \ldots + X_n$ and define $\mathcal{G}_n = \sigma(S_n, S_{n+1}, \ldots)$. Compute $E[X_1 \mid \mathcal{G}_n]$.

SOLUTION. It is easily checked that

$$\mathcal{G}_n = \sigma(S_n, S_{n+1}, \ldots) = \sigma(S_n, X_{n+1}, X_{n+2}, \ldots)$$

and, since X_1 is independent of $\sigma(X_{n+1}, X_{n+2}, \ldots)$, we have

$$E[X_1 \mid \mathcal{G}_n] = E[X_1 \mid \sigma(S_n)] = E[X_1 \mid S_n]$$

where we used the fact that if $\mathcal{H}$ is independent of $\sigma(\sigma(X), \mathcal{G})$, then $E[X \mid \sigma(\mathcal{G}, \mathcal{H})] = E[X \mid \mathcal{G}]$ a.s. In addition to this, as the X_i's are i.i.d. r.v.'s, it is also

$$E[X_i \mid S_n] = E[X_j \mid S_n] \;\; \forall i, j = 1, 2, \ldots$$

Now,

$$nE[X_1 \mid S_n] = \sum_{i=1}^{n} E[X_i \mid S_n] = E[S_n \mid S_n] = S_n$$

and therefore we have established that

$$E[X_1 \mid \mathcal{G}_n] = \frac{S_n}{n}.$$

PROBLEM 1.5.4. Let $X_1, X_2, \ldots$ be a sequence of independent r.v.'s each having zero mean and finite variance. Let $S_n = X_1 + X_2 + \ldots + X_n$ and $\mathcal{F}_n = \sigma(X_1, X_2, \ldots, X_n)$. Prove that

$$E[S_{n+1} \mid \mathcal{F}_n] = S_n \text{ and } \mathrm{Var}[S_{n+1} \mid \mathcal{F}_n] = \mathrm{Var}[X_{n+1}]$$

for all $n = 1, 2, \ldots$ Show also that

$$E[X_1 \mid \bar{X}] = \frac{S_n}{n}.$$

SOLUTION. For the first part, using the linearity property of conditional expectation, the fact that S_n is $\mathcal{F}_n$-measurable and X_{n+1} is independent of $\mathcal{F}_n$, we have

$$\begin{aligned} E[S_{n+1} \mid \mathcal{F}_n] &= E[S_n + X_{n+1} \mid \mathcal{F}_n] = E[S_n \mid \mathcal{F}_n] + E[X_{n+1} \mid \mathcal{F}_n] \\ &= S_n + E[X_{n+1}] = S_n + 0 = S_n. \end{aligned}$$

For the variance, using the fact that X_{n+1} is independent of $\mathcal{F}_n$ for all $n = 1, 2, \ldots$ (i.e. $E[\phi(X_{n+1})\theta(X_1, \ldots, X_n)) \mid \mathcal{F}_n] = E[\phi(X_{n+1})] \cdot \theta(X_1, \ldots, X_n)$ for any $\mathcal{F}_n$-measurable function $\theta(\cdot)$), we have

$$\begin{aligned} \mathrm{Var}[S_{n+1} \mid \mathcal{F}_n] &= E[S_{n+1}^2 \mid \mathcal{F}_n] - \left(E[S_{n+1} \mid \mathcal{F}_n]\right)^2 \\ &= E[(S_n + X_{n+1})^2 \mid \mathcal{F}_n] - \left(E[S_n + X_{n+1}) \mid \mathcal{F}_n]\right)^2 \\ &= E[S_n^2 + X_{n+1}^2 + 2S_nX_{n+1} \mid \mathcal{F}_n] - \left(E[S_n + X_{n+1}) \mid \mathcal{F}_n]\right)^2 \\ &= S_n^2 + E[X_{n+1}^2 \mid \mathcal{F}_n] + 2S_nE[X_{n+1} \mid \mathcal{F}_n] - (S_n + E[X_{n+1} \mid \mathcal{F}_n])^2 \\ &= S_n^2 + E[X_{n+1}^2 \mid \mathcal{F}_n] + 2S_nE[X_{n+1} \mid \mathcal{F}_n] \\ &- S_n^2 + (E[X_{n+1} \mid \mathcal{F}_n])^2 - 2S_nE[X_{n+1} \mid \mathcal{F}_n] \\ &= E[X_{n+1}^2 \mid \mathcal{F}_n] - (E[X_{n+1} \mid \mathcal{F}_n])^2 \\ &= E[X_{n+1}^2] - (E[X_{n+1}])^2 = \mathrm{Var}[X_{n+1}] \end{aligned}$$

as we were supposed to show.
It is easily verified that, because of the assumption that the X_i's are i.i.d., it must be $E[X_i \mid S_n] = E[X_j \mid S_n]$ $i, j = 1, 2, \ldots$ Then,

$$nE[X_1] = \sum_{k=1}^{n} E[X_k \mid S_n] = E[S_n \mid S_n] = S_n$$

using the linearity property of conditional expectation and the fact that S_n is $\sigma(S_n)$-measurable. Finally, it is also easy to show that

$$\sigma(\bar{X}) \subseteq \sigma(S_n).$$

Hence, using the Tower Property of conditional expectation, we find

$$E[X_1 \mid \bar{X}] = E[X_1 \mid \sigma(\bar{X})] = E[E[X_1 \mid \sigma(S_n)] \mid \sigma(\bar{X})]$$
$$= E[S_n/n \mid \sigma(\bar{X})] = E[\bar{X} \mid \sigma(\bar{X})] = \bar{X}.$$

PROBLEM 1.5.5. Let $X_1, X_2, \ldots, X_n$ be $\mathbf{R}$-valued random variables defined on some probability space $(\Omega, \mathcal{F}, P)$ which are mutually independent, have a continuous common distribution function, F and such that $E[X_1] < \infty$. Let $Y = \max\{X_1, X_2, \ldots, X_n\}$. Prove that for each fixed k, $1 \leq k \leq n$:

$$E[X_k \mid Y = y] = \frac{n-1}{n} \frac{1}{F(y)} \int_0^y s\, dF(s) + \frac{y}{n}.$$

SOLUTION. Fix k in $1, 2, \ldots, n$ and let $Q_{X|Y=y}(x) = Pr[X_k < x \mid Y = y]$. Then,

$$Q_{X|Y=y}(x) = \begin{cases} 1 & \text{if } x \geq y; \\ \lim_{h \downarrow 0} \frac{Pr[X_k < x, Y \in (y, y+h)]}{Pr[Y \in (y, y+h)]} & \text{if } x < y. \end{cases}$$

Now, using the assumption that the X_i's are mutually independent and have a continuous common distribution function, it is easy to show that when $x < y$,

$$Pr[X_k < x, Y \in (y, y+h)] = F(x) \sum_{j=0}^{n-2} \binom{n-1}{j} [F(y+h) - F(y)]^{n-1-j} F^j(y)$$

and, similarly, that

$$Pr[Y \in (y, y+h)] = \sum_{h=0}^{n-1} \binom{n}{h} [F(y+h) - F(y)]^{n-h} F^h(y).$$

Because of the assumption about the continuity of the distribution function F we have $A_h \equiv F(y+h) - F(y) \downarrow 0$ when $h \downarrow 0$. Now,

$$\begin{aligned}
&\lim_{h \downarrow 0} \frac{Pr[X_k < x, Y \in (y, y+h)]}{Pr[Y \in (y, y+h)]} \\
&= \lim_{h \downarrow 0} \frac{F(x) \sum_{j=0}^{n-2} \binom{n-1}{j} [A_h]^{n-1-j} F^j(y)}{\sum_{h=0}^{n-1} \binom{n}{h} [A_h]^{n-h} F^h(y)} \\
&= \lim_{h \downarrow 0} \frac{F(x)}{F(y)} \frac{\sum_{j=0}^{n-2} \binom{n-1}{j} [A_h]^{n-1-j} F^j(y)}{\sum_{h=0}^{n-1} \binom{n}{h} [A_h]^{n-h} F^{h-1}(y)} \\
&= \lim_{h \downarrow 0} \frac{F(x)}{F(y)} \frac{\sum_{j=0}^{n-2} \binom{n-1}{n-2-j} F^{n-2-j}(y) [A_h]^j}{\sum_{h=0}^{n-1} \binom{n}{n-1-h} F^{n-2-h}(y) [A_h]^h + A_h^{n-1}/F(y)} = \frac{(n-1)F(x)}{nF(y)}.
\end{aligned}$$

Therefore, we have found that

$$Q_{X|Y=y}(x) = \begin{cases} 1 & \text{if } x \geq y; \\ \frac{(n-1)F(x)}{nF(y)}. & \text{if } x < y. \end{cases}$$

For **R**-valued r.v.'s it is know that a regular conditional probability measure exists. In this case, the conditional expectation can be expressed as

$$E[X_k \mid Y = y](\omega) = \int_{-\infty}^{\infty} x \, dQ^Y(dx, \omega)$$

where $Q^Y(x, \cdot)$ is a regular conditional probability measures of $X \mid Y$ which is a version of $Pr[X \in dx \mid Y = y] = Q_{X|Y=y}(dx)$. Since $E[X]$ exists, so does $E[X \mid Y = y]$ and, in particular, we have

$$E[X \mid Y = y] = -\int_{-\infty}^{0} Q_{X|Y=y}[X < x] \, dx + \int_{0}^{y} Q_{X|Y=y}[X \geq x] \, dx$$

if $y \geq 0$ and

$$E[X \mid Y = y] = -\int_{-\infty}^{y} Q_{X|Y=y}[X < x] \, dx$$

when $y < 0$. In the case $y \geq 0$, replacing $Q_{X|Y=y}[X < x]$ with its expression above and using integration by parts yields

$$\begin{aligned}
E[X \mid Y = y] &= -\frac{n-1}{nF(y)} \int_{\infty}^{0} x \, dF(x) + \frac{1}{nF(y)} \int_{0}^{y} (nF(y) - (n-1)F(x)) \, dx \\
&= +\frac{n-1}{nF(y)} \int_{\infty}^{0} x \, dF(x) + \frac{1}{nF(y)} \Big[(nF(y) - (n-1)F(x))x \,|_0^y \\
&+ \frac{1}{nF(y)} \int_{0}^{y} (n-1)x \, dF(x) \Big] \\
&= \frac{y}{n} + \frac{n-1}{nF(y)} \int_{-\infty}^{y} x dF(x).
\end{aligned}$$

A similar result holds when $y < 0$ and the proof of the statement is complete.

PROBLEM 1.5.6. Let X and Y be random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. Prove that if

$$E[X \mid Y] = Y \text{ a.s. and } E[Y \mid X] = X \text{ a.s}$$

then $X = Y$ a.s. Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$ and assume that $E[X^2]$ exists. Prove that if X is a random variable on some probability space $(\Omega, \mathcal{F}, P)$, $\mathcal{G} \subset \mathcal{F}$ and $E[X^2] < \infty$ then

$$E[(X - E[X \mid \mathcal{G}])^2] = E[(X - E[X \mid \mathcal{F}])^2] + E[(E[X \mid \mathcal{F}] - E[X \mid \mathcal{G}])^2]$$

and show that

$$E[(X - E[X \mid \mathcal{F}])^2] \leq E[(X - E[X])^2].$$

SOLUTION. To prove the first statement, let's assume that $X \neq Y$ a.s. Then there exists $A \in \sigma(Y) \cap \sigma(X)$ such that $P(A) > 0$ and $X(\omega) \neq Y(\omega)$ for all $\omega \in A$. In particular, this implies that

$$\int_A X\, dP \neq \int_A Y\, dP \quad \text{for some } A \in \sigma(Y).$$

But, because of the assumption that $E[X \mid Y] = Y$ a.s. and using the definition of conditional expectation, we must have

$$\int_A X\, dP = \int_A E[X \mid Y]\, dP = \int_A Y\, dP \quad \forall A \in \sigma(Y).$$

The same reasoning applies if we assumed that $A \in \sigma(X)$. Hence, assuming that $X \neq Y$ a.s. leads to a contradiction and therefore it must be $X = Y$ a.s.

The conditions $E[X \mid Y] = Y$ a.s. and $E[Y \mid X] = X$ a.s. cannot be replaced by $E[X \mid Y] = Y$ a.s. or $[E[Y \mid X] = X$ a.s. unless it is known that either Y is $\sigma(X)$-measurable or X is $\sigma(Y)$-measurable. In general, this is not the case.

To prove the second statement we notice first that

$$\begin{aligned} E[(X - E[X \mid \mathcal{G}])^2] &= E[(X - E[X \mid \mathcal{F}] + E[X \mid \mathcal{F}] - E[X \mid \mathcal{G}])^2] \\ &= E[((X - E[X \mid \mathcal{F}]) + (E[X \mid \mathcal{F}] - E[X \mid \mathcal{G}]))^2] \\ &= E[(X - E[X \mid \mathcal{F}])^2] + E[(E[X \mid \mathcal{F}] - E[X \mid \mathcal{G}])^2] \\ &+ 2E[(X - E[X \mid \mathcal{F}])] \cdot (E[X \mid \mathcal{F}] - E[X \mid \mathcal{G}])]. \end{aligned}$$

Since for every random variable Z that is $\mathcal{H}$-measurable we have

$$E[Z] = E[E[Z \mid \mathcal{H}]]$$

we can use the fact for the random variable $(X - E[X \mid \mathcal{F}])(E[X \mid \mathcal{F}] - E[X \mid \mathcal{G}])$ which is clearly $\mathcal{F}$-measurable. Thus,

$$\begin{aligned} &E[(X - E[X \mid \mathcal{F}])(E[X \mid \mathcal{F}] - E[X \mid \mathcal{G}])] \\ &\quad = E[E[(X - E[X \mid \mathcal{F}])(E[X \mid \mathcal{F}] - E[X \mid \mathcal{G}]) \mid \mathcal{F}]] \\ &\quad = (E[X \mid \mathcal{F}] - E[X \mid \mathcal{G}]) \cdot E[E[(X - E[X \mid \mathcal{F}]) \mid \mathcal{F}] \\ &\quad = (E[X \mid \mathcal{F}] - E[X \mid \mathcal{G}]) \cdot E[E[X \mid \mathcal{F}] - E[X \mid \mathcal{F}]] = 0 \end{aligned}$$

and the proof of the second statement is therefore complete.
The last statement is easily shown to be correct by letting $\mathcal{G} = \{\emptyset, \Omega\}$. In fact, under this assumption $E[X \mid \mathcal{G}] = E[X]$, and since $E[(E[X \mid \mathcal{F}] - E[X \mid \mathcal{G}])^2] \geq 0$ the statement follows.

PROBLEM 1.5.7. Let X, Y be real random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. Assume that both X and Y are square-integrable and they are such that

$$E[X \mid Y] = Y, \quad E[Y \mid X] = 0.$$

Show that $Y = 0$ a.s.

SOLUTION. We start by showing that $E[XY] = 0$. In fact,

$$E[XY] = E[E[XY \mid X]] = E[X \cdot E[Y \mid X]] = E[X \cdot 0] = 0.$$

In addition, it is also:

$$Y^2 = Y \cdot E[X \mid Y]$$

and, taking expectation on both sides, we find

$$E[Y^2] = E[Y \cdot E[X \mid Y]] = E[XY] = 0$$

which clearly implies $Y = 0$ a.s.

PROBLEM 1.5.8. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $\mathcal{G}, \mathcal{H}$ be two sub-σ-algebras of $\mathcal{F}$. If X is a random variable such that $E[X \mid \mathcal{G}]$ exists, $\mathcal{G} \subset \mathcal{H}$, and $E[X \mid \mathcal{H}]$ is $\mathcal{G}$-measurable, then $E[X \mid \mathcal{G}] = E[X \mid \mathcal{H}]$.
Let $\mathcal{G}$ and $\mathcal{H}$ be two sub-σ-algebras of $\mathcal{F}$. Show that $E[X \mid \mathcal{H}] = E[X \mid \mathcal{G}]$ for every $X \in L^1((\Omega, \mathcal{F}, P))$, if and only if $\bar{\mathcal{G}} = \bar{\mathcal{H}}$[8]

SOLUTION. The first statement is proved by the following reasoning

$$E[X \mid \mathcal{G}] = E[E[X \mid \mathcal{G}] \mid \mathcal{H}]] = E[E[X \mid \mathcal{H}] \mid \mathcal{G}] = E[X \mid \mathcal{H}].$$

The first two equalities are justified by the tower property of conditional expectation, the third one by the fact that $E[X \mid \mathcal{H}]$ is $\mathcal{G}$-measurable, by assumption. To prove the second statement of the problem we observe that when we assume that $E[X \mid \mathcal{G}] = E[X \mid \mathcal{H}]$ we are assuming that $E[X \mid \mathcal{G}]$ has an $\mathcal{H}$-measurable version. So, if we take $G \in \mathcal{G}$ and let $X = I_G$, we have in first place that $E[I_G \mid \mathcal{G}] = I_G$ and, in addition, there is a $\mathcal{H}$-measurable version for it, Y_G. Define $A = \{\omega \in \Omega : Y_G(\omega) = 1\}$. Clearly, $A \in \mathcal{H}$ as, by assumption, Y_G is the $\mathcal{H}$ measurable version of $E[I_G \mid \mathcal{G}]$. Then, $G \Delta A \subset \{I_G \neq Y_G\} \in \mathcal{N}$. According to the definition of $\bar{\mathcal{H}}$ this means that $G \in \bar{\mathcal{H}}$ and, since G is a generic element of $\mathcal{G}$, it is established that $\mathcal{G} \subset \bar{\mathcal{H}}$. In turn, this implies that $\bar{\mathcal{G}} \subset \bar{\mathcal{H}}$. In fact, adding P-null sets to the sets in $\mathcal{G}$ gives new subsets of Ω which still differ from sets of $\mathcal{H}$ by P-null sets. The argument can obviously be reversed to show that $\bar{\mathcal{H}} \subset \bar{\mathcal{G}}$ and, therefore, $\bar{\mathcal{G}} = \bar{\mathcal{H}}$. If, on the other hand, we assume that $\bar{\mathcal{H}} = \bar{\mathcal{G}}$, it is clear that $E[X \mid \mathcal{G}] = E[X \mid \mathcal{H}]$. In fact, that would mean that the elements of $\mathcal{G}$ and $\mathcal{H}$ differ by P-null sets only and P-null sets do not effect the value of conditional expectations.

1.6 Jensen's Inequality, Hölder Inequality

PROBLEM 1.6.1. Suppose we observe a single bivariate random vector (W, Y) where

$$W = (1 + T)X \text{ and } Y = \theta X + V.$$

[8] For any σ-algebra, $\mathcal{A}$, of $(\Omega, \mathcal{F}, P)$, $\bar{\mathcal{A}}$ denotes the augmentation of $\mathcal{A}$ defined as $\bar{\mathcal{A}} = \sigma(\mathcal{A} \cup \mathcal{N})$ where $\mathcal{N}$ is the collection of all P-null subsets of Ω. Equivalently, one can show that $\bar{\mathcal{A}} = \{S \in \Omega : \exists A \in \mathcal{A} \text{ and } S \Delta A \in \mathcal{N}\}$.

The variables T, X, and V are independent. Assume that T is symmetric about the origin and concentrated on $[-0.5, 0.5]$; $E[V] = 0$; $X \neq 0$; and θ is a positive number. Define a new random variable (estimator) $\hat{\theta} = Y/W$ and a function (bias) $b(\theta) = E[\hat{\theta}] - \theta$. Show

(a) $b(\theta)/\theta$ does not depend on θ,

(b) $0 \leq b(\theta)/\theta \leq 1/3$.

SOLUTION. The random variable $\hat{\theta}$ can be written as

$$\hat{\theta} = \frac{Y}{W} = \frac{\theta X + V}{(1+T)X} = \frac{\theta}{1+T} + V\frac{1}{1+T}\frac{1}{X}.$$

Then,

i. $\hat{\theta}$ is well defined as $X \neq 0$ and $(1+T)^{-1} \neq 0$ because of the assumption $T \in [-0.5, 0.5]$;

ii. T, X, V are independent by assumption and this implies that

$$E[g(T)h(X)k(V)] = E[g(T)]E[h(X)]E[k(V)]$$

for arbitrary measurable functions g, h and k. Therefore:

$$E[V \cdot (1+T)^{-1} \cdot X^{-1}] = E[V] \cdot E[(1+T)^{-1}] \cdot E[X^{-1}];$$

iii. $E[V] = 0, E[(1+T)^{-1}] \in [2/3, 2]$ and assume $E[X^{-1}]$ exists;

iv. this gives $b(\theta) \equiv E[\hat{\theta}] - \theta = \theta[E[(1+T)^{-1}] - 1]$ and hence $b(\theta)/\theta = E[(1+T)^{-1}] - 1$ and as T doesn't depend on θ, $b(\theta)/\theta$ doesn't either. This completes the proof of (a).

There are at least two different ways to handle part (b).
Method 1: Jensen's Inequality. It is easily checked that $g(t) = -t/(1+t)$ is convex on $[-0.5, 0.5]$ and hence by Jensen's Inequality $b(\theta)/\theta = E[g(T)] \geq 0$. Also, the convexity of $g(t)$ entitles us to say that $g(t)$ is less than or equal to the convex combination of the function evaluated at the endpoints of the interval, that is, $g(t) \leq (1-4t)/3$ and $b(\theta)/\theta = E[g(t)] \leq (1 - 4E[T])/3 = 1/3$.
Method 2: Taylor Expansion and DCT. A different way to compute the two bounds for $b(\theta)/\theta$ is to take the Taylor's expansion of $1/(1+T)$ about 0, i.e.:

$$\frac{1}{1+T} = 1 + \sum_{n=1}^{\infty}(-1)^n T^n$$

so that

$$E[\frac{1}{1+T}] = 1 + E[\sum_{n=0}^{\infty}(-1)^n T^n].$$

If we define $S_n = \sum_{m=0}^{n}(-1)^m T^m$, it is easily seen that $S_n \leq 2 \equiv S\ \forall T \in [-0.5, 0.5]$ and as a degenerate r.v. S has evidently finite expectation. In addition, we have we that $\sum_{n=1}^{m}(-1)^n T^n \to 1/(1+T)$ for every choice of $T \in [-0.5, 0.5]$ and hence the convergence is with probability 1. As all the assumptions to use the DCT for infinite sums of random variables are satisfied we have now:

$$E[\sum_{n=0}^{m}(-1)^n T^n] = 1 + \sum_{n=1}^{m}(-1)^n E[T^n] \to E[1/(1+T)]$$

when $m \to \infty$. It is easy to verify that $E[T^n] = 0$ if n is odd, while $E[T^n] \leq (.5)^n$ when n is an even number. Therefore,

$$1 \leq E[1/(1+T)] \leq 1 + \sum_{n=1}^{\infty}(0.5)^2 n = 1 + 1/3$$

and we have again that

$$0 = 1 - 1 = b(\theta)/\theta \leq 1 + 1/3 - 1 = 1/3.$$

PROBLEM 1.6.2. Let $1 \leq p \leq r < \infty$ and consider $Y \in \mathcal{L}^r$. Prove that $||Y||_p \leq ||Y||_r$.

SOLUTION. For every $n \in \mathbf{N}$, define

$$X_n(\omega) = [| Y(\omega) | \wedge n]^p \quad \forall \omega \in \Omega.$$

Then, X_n is bounded and thus, X_n and $X_n^{r/p} \in \mathcal{L}^1$.
Let now, $f(x) = x^{r/p} I_{[0,\infty)}(x)$. It is easily seen that this is a convex function, hence using Jensen's Inequality one gets

$$(E[X_n])^{r/p} \leq E[X_n^{r/p}] = E[(| Y | \wedge n)^r] \leq E[| Y |^r].$$

Clearly, $X_n \to | Y |^p$ as $n \to \infty$ and, since the X_n's form an increasing sequence of positive random variables, we can use the MCT. This gives

$$E[X_n] \nearrow E[| Y |^p$$

and the previous inequality can now be rewritten as

$$(E[| Y |^p])^{r/p} \leq E[| Y |^r].$$

Raising both terms to the power $1/r$, gives

$$((E[| Y |^p])^{r/p})^{1/r} \leq (E[| Y |^r])^{1/r}$$
$$(E[| Y |^p])^{1/p} \leq (E[| Y |^r])^{1/r}$$

or

$$|| Y ||_p \leq || Y ||_r$$

as we were supposed to prove.

PROBLEM 1.6.3. Prove Hölder's Inequality:

$$E[| XY |] \leq E^{1/p}[| X |^p] \cdot E^{1/q}[| Y |^q]$$

where $p^{-1} + q^{-1} = 1$.

(a) First show that $E[X^{1/p}Y^{1/q}] \leq E^{1/p}[X^p] \cdot E^{1/q}[Y^q]$ for positive random variables X and Y with $p^{-1} + q^{-1} = 1$ by using Jensen's Inequality.

(b) Then deduce Hölder's Inequality from part (a).

SOLUTION. Let $f(X,Y) = -X^{1/p} \cdot Y^{1/q}$ where X and Y are nonnegative variables and $p^{-1} + q^{-1} = 1$. Then:

Proposition. *$f(X,Y)$ as defined above is a convex function.*

Proof. To prove this fact it suffices to show that the Hessian, $H[f(X,Y)]$ is a positive semidefinite matrix. In this case the Hessian matrix turns out to be:

$$H[f(X,Y)] = \begin{pmatrix} \frac{1}{pq}x^{1/p-2}y^{1/q} & -\frac{1}{pq}x^{-1/q}y^{-1/p} \\ -\frac{1}{pq}x^{1/q}y^{-1/p} & \frac{1}{pq}x^{1/p}y^{1/q-2} \end{pmatrix}$$

There are several ways to prove that this last matrix is positive semidefinite. Since we have a 2×2 matrix we can use the definition and show that $\forall (t_1, t_2) \in \mathbf{R}^2$ such that $(t_1, t_2) \neq (0,0)$

$$(t_1, t_2) \cdot H[f(X,Y)](t_1, t_2)^T \geq 0.$$

In our case, we have:

$$\begin{aligned}(t_1, t_2) \cdot H[f(X,Y)](t_1, t_2)^T &= t_1^2 x^{1/p-2}y^{1/q} + t_2^2 x^{1/p}y^{1/q-2} - 2t_1t_2x^{1/p-1}y^{1/q-1} \\ &= (t_1 x^{1/(2p)-1}y^{1/(2q)} - t_2 x^{1/(2p)}y^{1/(2q)-1})^2 \geq 0\end{aligned}$$

for every choice of $(t_1, t_2) \neq (0,0)$.

□

Then, in the case X and Y are positive r.v.'s, the bivariate Jensen's Inequality gives

$$E[-X^{1/p}Y^{1/q}] > -(E[X])^{1/p}(E[Y])^{1/q}$$

$$E[X^{1/p}Y^{1/q}] < (E[X])^{1/p}(E[Y])^{1/q}.$$

Replacing X and Y by $| X |^p$ and $| Y |^q$, respectively, gives the desired inequality.

PROBLEM 1.6.4. Suppose that X and Y are nonnegative random variables, $r > 1$, and

$$P[Y \geq t] \leq \frac{1}{t} \int_{Y \geq t} X dP$$

for $t > 0$. Use Fubini's Theorem and Hölder's Inequality to prove

$$E[Y^r] \leq \left(\frac{r}{r-1}\right)^r E[X^r].$$

SOLUTION. According to the definition and using the fact that by assumption Y is a nonnegative r.v., we have

$$E[Y^r] = \int_0^\infty y^r dF(y)$$

which can also be written as

$$E[Y^r] = \int_0^\infty \left(\int_0^y rz^{r-1} dz\right) dF(y).$$

As we are dealing with nonnegative quantities we can use Fubini's Theorem to interchange the order of integration. This gives

$$E[Y^r] = \int_0^\infty \left(\int_z^\infty dF(y)\right) rz^{r-1} dz = \int_0^\infty P[Y \geq z] rz^{r-1} dz.$$

Using the inequality given in the text and Hölder's Inequality we can also write

$$\begin{aligned} E[Y^r] &\leq \int_0^\infty \int_{Y \geq z} \frac{1}{z} X dP rz^{r-1} dz = \int_0^\infty \int_{Y \geq z} X dP rz^{r-2} dz \\ &= \frac{r}{r-1} \int_0^\infty XY^{r-1} dP = \frac{r}{r-1} E[XY^{r-1}] \leq \frac{r}{r-1} E^{1/r}[X^r] E^{(r-1)/r}[Y^r] \end{aligned}$$

and raising both sides to the r–th power gives

$$E^r[Y^r] \leq \left(\frac{r}{r-1}\right)^r E[X^r] E^{r-1}[Y^r]$$

and, dividing both sides by $E^{r-1}[Y^r]$,

$$E[Y^r] \leq \left(\frac{r}{r-1}\right)^r E[X^r]$$

as we were supposed to prove.

1.7 Markov's Inequality, Cauchy-Schwarz Inequality

PROBLEM 1.7.1. Given p, $Y_1, Y_2, \ldots$ are distributed i.i.d. Bernoulli(p). Suppose p is random, with distribution μ and assume that μ is not concentrated at $\{p_0\}$ for any $p_0 \in [0,1]$, nor on $\{0,1\}$. Define $S_n = Y_1 + Y_2 + \ldots + Y_n$, for $n = 1, 2, \ldots$

(a) Prove that $p(Y_2 = 1 \mid Y_1 = 1) > p(Y_2 = 1 \mid Y_1 = 0)$;

(b) Prove that for $n \geq 1$, the function of k

$$g_n(k) = p(Y_{n+1} = 1 \mid S_n = k)$$

is strictly increasing on its domain $\{0, 1, \ldots, n\}$.

SOLUTION. It is easily seen that we can write

$$p(Y_1 = y_1, Y_2 = y_2) = \int_0^1 p^{y_1+y_2}(1-p)^{2-y_1-y_2}\mu(dp)$$

so that summing out Y_2, gives

$$p(Y_1 = y_1) = \int_0^1 p^{y_1}(1-p)^{2-y_1}\mu(dp) + \int_0^1 p^{y_1+1}(1-p)^{1-y_1}\mu(dp).$$

Therefore, using these formulae above, we can compute $p(Y_2 = 1 \mid Y_1 = 1)$ as

$$\begin{aligned} p(Y_2 = 1 \mid Y_1 = 1) &= \frac{p(Y_1 = 1, Y_2 = 1)}{p(Y_1 = 1)} \\ &= \frac{\int_0^1 p^2\mu(dp)}{\int_0^1 p(1-p)\mu(dp) + \int_0^1 p^2\mu(dp)} \end{aligned}$$

and, similarly, we find that

$$\begin{aligned} p(Y_2 = 1 \mid Y_1 = 0) &= \frac{p(Y_1 = 0, Y_2 = 1)}{p(Y_1 = 0)} \\ &= \frac{\int_0^1 p(1-p)\mu(dp)}{\int_0^1 p(1-p)\mu(dp) + \int_0^1 (1-p)^2\mu(dp)}. \end{aligned}$$

To show that the first inequality holds it is now enough to check that the following inequality holds:

$$\frac{\int_0^1 p^2\mu(dp)}{\int_0^1 p(1-p)\mu(dp) + \int_0^1 p^2\mu(dp)} > \frac{\int_0^1 p(1-p)\mu(dp)}{\int_0^1 p(1-p)\mu(dp) + \int_0^1 (1-p)^2\mu(dp)}$$

or, after simple manipulations, that

$$\int_0^1 p^2\mu(dp)\int_0^1 p(1-p)\mu(dp) > \Big(\int_0^1 p(1-p)\mu(dp)\Big)^2$$

which can be rewritten as

$$E[p^2]E[(1-p)^2] > (E[p(1-p)])^2.$$

It is now easily seen that last inequality is nothing but the Cauchy Schwarz Inequality. In general, we know that this inequality can be an equality iff $p = k(1-p)$. But, this would imply that $p = k/(1+k)$ and as by assumption the distribution μ is not concentrated at any of the points in $[0,1]$, we know that $p = k/(1-k)$ with probability zero so the inequality above is strict with probability one.

For the second part it is important to notice that Y_{n+1} and S_n are independent as S_n is a measurable function of $(Y_1, Y_2, \dots Y_n)$ and therefore the same procedure used for part (a) yields

$$p(Y_{n+1} = y_{n+1}, S_n = s) = \int_0^1 \binom{n}{s} p^{s+y_{n+1}}(1-p)^{n+1-s-y_{n+1}}\mu(dp)$$

where $y_{n+1} = 0, 1$ and $s = 0, 1, \dots, n$. Thus,

$$p(S_n = s) = \int_0^1 \binom{n}{s} p^s(1-p)^{n-s+1}\mu(dp) + \int_0^1 \binom{n}{s} p^{s+1}(1-p)^{n-s}\mu(dp).$$

Using these formulae we get that

$$p(Y_{n+1} = 1 \mid S_n = k) = \frac{\int_0^1 p^{k+1}(1-p)^{n-k}\mu(dp)}{\int_0^1 p^{k+1}(1-p)^{n-k}\mu(dp) + \int_0^1 p^k(1-p)^{n-k+1}\mu(dp)}$$

and, similarly,

$$p(Y_{n+1} = 1 \mid S_n = k-1) = \frac{\int_0^1 p^k(1-p)^{n-k+1}\mu(dp)}{\int_0^1 p^k(1-p)^{n-k+1}\mu(dp) + \int_0^1 p^{k-1}(1-p)^{n-k+2}\mu(dp)}.$$

Then, proving that the inequality in the text holds is the same as proving that the following inequality holds:

$$\int_0^1 p^{k+1}(1-p)^{n-k}\mu(dp) \int_0^1 p^{k-1}(1-p)^{n-k+2}\mu(dp) > \Big(\int_0^1 p^k(1-p)^{n-k+1}\mu(dp)\Big)^2$$

or that

$$\begin{aligned} &E[(p^{(k+1)/2}(1-p)^{(n-k)/2})^2]E[(p^{(k-1)/2}(1-p)^{(n-k+2)/2})^2] \\ &> (E[p^{(k+1)/2}(1-p)^{(n-k)/2}p^{(k-1)/2}(1-p)^{(n-k+2)/2}])^2 \\ &= (E[p^k(1-p)^{n-k+1}])^2. \end{aligned}$$

Again, the Cauchy Schwarz Inequality gives the desired strict inequality.

PROBLEM 1.7.2. Assume that a sequence of random variables, $\{X_t : t \in [0,1]\}$ defined on some probability space $(\Omega, \mathcal{F}, P)$ is such that there exist $p \geq 0$, $q \geq 0$, $C \geq 0$, and $r > 0$ so that for each $0 \leq t_1 < t_2 < t_3 \leq 1$:

$$E[|X_{t_1} - X_{t_2}|^p \cdot |X_{t_2} - X_{t_3}|^q] \leq C|t_3 - t_1|^{1+r}.$$

Prove that

$$P(\{\omega : |X_{t_1}(\omega) - X_{t_2}(\omega)| \geq \epsilon_1\} \cap \{\omega : |X_{t_2}(\omega) - X_{t_3}(\omega)| \geq \epsilon_2\}) \leq E\left[\frac{|X_{t_1} - X_{t_2}|^p \cdot |X_{t_2} - X_{t_3}|^q}{\epsilon_1^p \epsilon_2^q}\right] \leq \frac{C|t_3 - t_1|^{1+r}}{\epsilon_1^p \epsilon_2^q}.$$

SOLUTION. All there is to prove is

$$Pr[\{\omega : |X_{t_1}(\omega) - X_{t_2}(\omega)| \geq \epsilon_1\} \cap \{\omega : |X_{t_2}(\omega) - X_{t_3}(\omega)| \geq \epsilon_2\}] \leq E\left[\frac{|X_{t_1} - X_{t_2}|^p \cdot |X_{t_2} - X_{t_3}|^q}{\epsilon_1^p \epsilon_2^q}\right]$$

and this is easy to establish since

$$\begin{aligned} E[|X|^p|Y|^q] &\geq \int_{|X|\geq\epsilon_1, |Y|\geq\epsilon_2} |X|^p|Y|^q \, dP_{XY} \\ &\geq \epsilon_1^p \epsilon_2^q \int_{|X|\geq\epsilon_1, |Y|\geq\epsilon_2} dP_{XY} = \epsilon_1^p \epsilon_2^q P(|X| \geq \epsilon_1, |Y| > \epsilon_2). \end{aligned}$$

PROBLEM 1.7.3. Let $X_1, X_2, \ldots,$ be independent, identically distributed random variables with finite second moment. Prove that

(a) $nP[| X_1 | \geq \epsilon\sqrt{n}] \to 0$;

(b) $(1 - P[| X_1 | \geq \epsilon\sqrt{n}])^n \to 1$; and

(c) $n^{-1/2} \max_{k \leq n} | X_k | \xrightarrow{P} 0$ as $n \to \infty$.

SOLUTION. Since by assumption $X_1, X_2, \ldots,$ have finite second moment, we can use Markov's Inequality and write

$$P[| X | \geq \alpha] \leq \frac{1}{\alpha^2} \int_{|X|\geq\alpha} | X |^2 \, dP \leq \frac{1}{\alpha^2} E[| X |^2].$$

Letting now $X = X_1$ and $a = \epsilon \cdot n^{1/2}$, we find that for any $\epsilon > 0$, it is

$$\begin{aligned} nP[| X_1 | \geq \epsilon\sqrt{n}] &\leq n \frac{1}{n\epsilon^2} \int_{|X_1|^2 \geq \epsilon\sqrt{n}} | X |^2 \, dP \\ &= \frac{1}{\epsilon} \int_{|X|\geq\epsilon\sqrt{n}} | X_1 |^2 \, dP \to 0 \text{ as } n \to \infty. \end{aligned}$$

From (a) we have also

$$P[| X_1 | \geq \epsilon\sqrt{n}] = O(n^{-\alpha})$$

for some $\alpha > 1$ and it is possible to use this fact to prove (b). In fact, it is true in general that

$$\lim_{n\to\infty}\left(1-\frac{1}{n^\alpha}\right)^n \to 1$$

as $n \to \infty$ if $\alpha > 1$. To prove this fact it suffices to apply De L'Hôpital's Rule once.

To prove (c) it is enough to notice that for any arbitrary positive constant ϵ, it is

$$\begin{aligned} P[n^{-1/2}\max_{k\leq n} \mid X_k \mid \leq \epsilon] &= \Pi_{k=1}^n P[n^{-1/2} \mid X_k \mid \leq \epsilon] \\ &= (1 - P[\mid X_1 \mid \geq \epsilon\sqrt{n}])^n \to 1 \end{aligned}$$

as $n \to \infty$ because of what proved in (b). As ϵ is arbitrary, this proves that

$$n^{-1/2}\max_{k\leq n} \mid X_k \mid \xrightarrow{P} 0$$

as $n \to \infty$.

PROBLEM 1.7.4. Give a sense in which Chebyshev's Inequality is best possible.

SOLUTION. The possibility to transform Chebyshev's Inequality into an equality relies on the following lemma (*Ghosh and Meeden, The American Statistician; 1977, pp. 35-6).*

Lemma. *Let Y be a random variable with $P[Y \geq 0] = 1$ and $P[Y = 0] < 1$. Then $\forall a > 0$*

$$P[Y \geq a] \leq E[Y]/a$$

equality holding iff $P[Y = a] = p = 1 - P[Y = 0]$, $p \in (0, 1]$.

Proof.

$$E[Y] = E[Y \cdot I_{[Y<a]}] + E[Y \cdot I_{[Y\geq a]}] \geq E[Y \cdot I_{[Y<a]}] + aP[Y \geq a] \geq aP[Y \geq a]$$

equality holding iff $E[Y \cdot I_{[Y<a]}] = 0$, $P[Y \geq a] > 0$ and $P[Y = a] \neq 0$. This is however equivalent to $P[Y = l] = 0\ \forall l \in \mathbf{R}^+ - \{a\}$ and $P[Y = a] \neq 0$. $\square$

Now, assume that X is a random variable with mean μ and variance σ^2. Then, if $k > 0$, it is

$$P[\mid X - \mu \mid \geq k\sigma] = P[(X - \mu)^2 \geq k^2\sigma^2]$$

and, letting $Y \equiv (X - \mu)^2$, $a \equiv k^2\sigma^2$, the equality in the lemma stated above is attained iff

$$P[(X - \mu)^2 = k^2\sigma^2] = p = 1 - P[(X - \mu)^2 = 0].$$

It is easily seen that this is possible iff X have mass on at most three points. So, we can conclude that Chebyshev's Inequality is best possible when the random variable X is of the type[9]

$$X = \begin{cases} \mu + a & \text{with probability } p/2, \\ \mu - a & \text{with probability} p/2, \\ \mu & \text{with probability } 1 - p. \end{cases}$$

with $a \geq k\sigma$. In this case, the Chebyshev's Inequality becomes

$$P[\mid X - \mu \mid \geq k\sigma] = p.$$

PROBLEM 1.7.5. Suppose X is a p dimensional multivariate normal vector with mean 0 and covariance matrix Σ. Let R be the region

$$\{X : tr(A\Sigma) - 2\sqrt{tr(A\Sigma)} \leq X'AX \leq tr(A\Sigma) + 2\sqrt{tr(A\Sigma)}\}$$

where A is a positive definite symmetric matrix. Show that

(a)
$$Pr[X \in R] \geq 1 - \frac{Var(X'AX)}{4tr(A\Sigma)};$$

(b) $Pr[X \in R] \geq 1/2$ when $A = \Sigma^{-1}$.

SOLUTION. The first part is a simple application of Chebyshev's Inequality to the quadratic form $Y = X'AX$. In fact,

$$E[X'AX] = E[tr(XX'A)] = tr[(\Sigma + \mu\mu')A] = tr(\Sigma A) + tr(\mu' A\mu)$$

and in our special case where $\mu = 0$ this formula gives:

$$E[X'AX] = tr(\Sigma A).$$

Now, using Chebyshev's Inequality we find

$$Pr[\mid X'AX - tr(A\Sigma) \mid \geq 2\sqrt{tr(A\Sigma)}] \leq 1 - \frac{Var(X'AX)}{4tr(A\Sigma)}$$

as we were suppose to prove.

Let now $A = \Sigma^{-1}$, then it is possible to write $X'AX$ as $X'\Sigma^{-1}X = X'\Sigma^{-1/2}\Sigma^{-1/2}X$. If we let $Z = \Sigma^{-1/2}X$, it follows that $Z \sim N_p(0, I)$ and therefore

$$X'\Sigma^{-1}X = \sum_{i=1}^{p} Z_i^2 \sim \chi^2_{(p)}.$$

[9]In this case $E[X] = \mu$ and $Var[X] = a^2 p$.

Thus, $Var(X'\Sigma^{-1}X) = 2p$ and $tr(A\Sigma)= tr(\Sigma^{-1}\Sigma) = tr(I_p) = p$. Replacing $Var(X'AX)$ and $tr(A\Sigma)$ in the formula above with these two new expressions gives

$$Pr[X \in R] \geq 1 - \frac{2p}{4p} = \frac{1}{2}.$$

PROBLEM 1.7.6. Let ξ and η be random variables with correlation coefficient ρ. Establish the following two-dimensional analog of Chebyshev's Inequality:

$$P[\mid \xi - E\xi \mid > \epsilon\sqrt{V\xi} \text{ or } \mid \eta - E\eta \mid > \epsilon\sqrt{V\eta}] \leq \epsilon^{-2}(1 + \sqrt{1-\rho^2})$$

SOLUTION. In order to prove the statement above we need the following

Lemma. *Let ξ and η be random variables with $E\xi = E\eta = 0$, $V\xi = V\eta = 1$ and correlation coefficient ρ. Then,*

$$E[\max\{\xi^2, \eta^2\}] \leq 1 + \sqrt{1-\rho^2}.$$

Proof. It is easily seen that $\max\{\xi^2, \eta^2\}$ can be written as

$$\max\{\xi^2, \eta^2\} = \frac{\mid \xi^2 + \eta^2 \mid + \mid \xi^2 - \eta^2 \mid}{2} =$$

$$= \frac{\xi^2 + \eta^2 + \mid \xi^2 - \eta^2 \mid}{2}.$$

This implies also that

$$E[\max\{\xi^2, \eta^2\}] = \frac{E[\xi^2] + E[\eta^2] + E[\mid \xi^2 - \eta^2 \mid]}{2}.$$

By the Cauchy-Schwarz Inequality we know that it must be

$$E[\mid \xi^2 - \eta^2 \mid] = E[\mid (\xi + \eta)(\xi - \eta) \mid] \leq$$

$$E^{1/2}[(\xi + \eta)^2]E^{1/2}[(\xi - \eta)^2] = [2(1+\rho)]^{1/2}[2(1-\rho)]^{1/2} = 2\sqrt{1-\rho^2}.$$

and finally, since by assumption $E\xi = E\eta = 0$ and $V\xi = V\eta = 1$, we can write

$$E[\max\{\xi^2, \eta^2\}] \leq \frac{1 + 1 + 2\sqrt{1-\rho^2}}{2} = 1 + \sqrt{1-\rho^2}.$$

□

Rewriting

$$P[\mid \xi - E\xi \mid > \epsilon\sqrt{V\xi} \text{ or } \mid \eta - E\eta \mid > \epsilon\sqrt{V\eta}]$$

as

$$P[|\ (\xi - E\xi)\ |\ /\sqrt{V\xi} > \epsilon \text{ or } |\ (\eta - E\eta)\ |\ /\sqrt{V\eta} > \epsilon] \equiv P[|\ \alpha\ |> \epsilon \text{ or } |\ \beta\ |> \epsilon]$$

where $\alpha \equiv (\xi - E\xi)/\sqrt{V\xi}$ and $\beta \equiv (\eta - E\eta)/\sqrt{V\eta}$, it is easily checked that α and β satisfy the assumptions of the Lemma above. Thus, it is possible to write[10]

$$P[|\ \alpha\ |> \epsilon \text{ or } |\ \beta\ |> \epsilon] = P[\max\{|\ \alpha\ |, |\ \beta\ |\} > \epsilon]$$
$$= P[\max\{\alpha^2, \beta^2\} > \epsilon^2] \leq \epsilon^{-2} E[\max\{\alpha^2, \beta^2\}] \leq \frac{1 + \sqrt{1-\rho^2}}{\epsilon^2}$$

using the univariate Chebyshev's Inequality and the Lemma above, respectively.

1.8 Stein's Lemma Analogues

PROBLEM 1.8.1. Prove the following analogues to Stein's Lemma, assuming appropriate conditions on the function g.

(a) If $X \sim \text{beta}(\alpha, \beta)$, then

$$E\Big[g(x)\Big(\beta - (\alpha - 1) \cdot \frac{1-X}{X}\Big)\Big] = E[(1-X) \cdot g'(X)].$$

(b) If $X \sim \text{gamma}(\alpha, \beta)$, then

$$E(g(x) \cdot (X - \alpha\beta) = \beta E[X \cdot g'(x)].$$

SOLUTION. In this problem all we need is integration by parts and an appropriate assumption about the function $g(\cdot)$. In the first case we have that

$$E\Big[g(x)\Big(\beta - (\alpha - 1) \cdot \frac{1-X}{X}\Big)\Big]$$
$$= \beta E[g(x)] + \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} (\alpha - 1) x^{\alpha-2} g(x) (1-x)^{\beta} dx.$$

[10] In fact, $P[T > \epsilon \text{ or } S > \epsilon]$ is the same as $1 - P[T \leq \epsilon \text{ and } S \leq \epsilon] = 1 - P[\max\{T, S\} \leq \epsilon] = P[\max\{T, S\} > \epsilon$.]

If one thinks of $(\alpha-1)x^{\alpha-2}$ as du and of $g(x)(1-x)^{\beta}$ as v, then we find that

$$\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}(\alpha-1)x^{\alpha-2}g(x)(1-x)^{\beta}dx = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}g(x)x^{\alpha-1}(1-x)^{\beta}|_0^1$$
$$-\int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}g'(x)x^{\alpha-1}(1-x)^{\beta}dx - \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\beta g(x)x^{\alpha-1}(1-x)^{\beta-1}dx$$
$$= 0 + \int_0^1 [(1-x)g'(x)]\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}dx$$
$$-\beta\int_0^1 g(x)\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}dx$$
$$= \beta E[g(X)] + E[(1-X)g'(X)] - \beta E[g(X)] = E[(1-X)g'(X)]$$

which holds for every differentiable function g such that $E[\mid g'(X)\mid] < \infty$.
The proof of (b) is similar to the previous one. In fact, it is easily seen that

$$E[g(X)(X-\alpha\beta)] = \int_0^{\infty} g(x)\frac{1}{\Gamma(\alpha)\beta^{\alpha}}x^{\alpha}e^{-x/\beta}dx - \alpha\beta E[g(X)].$$

Let now $g(x)x^{\alpha} = u$ and $(1/\beta)e^{-x/\beta} = dv$, then we have

$$E[g(X)(X-\alpha\beta)] = -\beta\frac{1}{\Gamma(\alpha)\beta^{\alpha}}x^{\alpha}e^{-x/\beta}|_0^{\infty} + \beta\int_0^{\infty} g'(x)\frac{1}{\Gamma(\alpha)\beta^{\alpha}}x^{\alpha}e^{-x/\beta}dx$$
$$+\beta\int_0^{\infty} g(x)\frac{1}{\Gamma(\alpha)\beta^{\alpha}}\alpha x^{\alpha-1}e^{-x/\beta}dx - \alpha\beta E[g(X)]$$
$$= 0 + \beta E[X\cdot g'(X)] + \alpha\beta E[g(X)] - \alpha\beta E[g(X)] = \beta E[X\cdot g'(X)]$$

which holds for every differentiable function g such that $E[\mid g'(X)\mid] < \infty$. This last formula is very useful to a quick computation of $Var[X]$. In fact, it suffices to take $g(X) = X - \alpha\beta$ to find that $Var[X] = \alpha\beta^2$.

1.9 Probability Inequalities, Bernstein's Inequality

PROBLEM 1.9.1. Let X be an $\mathbf{R}$-valued random variable defined on $(\Omega, \mathcal{F}, P)$ with finite, nonzero, second moment and nonnegative first moment. Prove that for $0 < \lambda < 1$

$$P[\{\omega : X(\omega) > \lambda E[X]\}] \geq \frac{(1-\lambda)^2(E[X])^2}{E[X^2]}.$$

Then, if $E[\mid X\mid] = 1$, show that this result provides a lower bound for the Chebyshev's Inequality.

SOLUTION. Let $B = \{\omega \in \Omega : X(\omega) > \lambda E[X]\}]$ and consider the new random variable $Y(\omega) = X\cdot I_B(\omega)$, with I_B the indicator function of the event B. Then, it is clearly

$$Y(\omega) = X\cdot I_B(\omega) = X(\omega) - X\cdot I_{B^c}(\omega)$$

and, using the properties of expectation, we find that

$$E[Y] = E[X] - E[X \cdot I_{B^c}] \geq E[X] - E[\lambda E[X]] = E[X] - \lambda E[X] = (1-\lambda)E[X].$$

Now, as by assumption it is $\lambda \in (0,1)$ and $E[X] \geq 0$, we have also

$$(E[Y])^2 \geq (1-\lambda)^2 (E[X])^2.$$

Thus, we find that

$$(E[Y])^2 = (E[X \cdot I_B])^2 \leq E[X^2] \cdot E[I_B^2] = E[X^2] \cdot E[I_B] = E[X^2] \cdot P[B]$$

where this last inequality is just the consequence of having applied the Cauchy-Schwartz Inequality, which we can do because of the assumption that $E[X^2] < \infty$ and, of course, because $E[I_B] < \infty$. Combining the last two inequalities together gives

$$P[B]E[X^2] \geq (E[Y])^2 \geq (1-\lambda)^2 (E[X])^2$$

and therefore

$$P[B] \geq \frac{(1-\lambda)^2 (E[X])^2}{E[X^2]}$$

as we were asked to prove. This last expression is well defined as, by assumption, we know that $0 < E[X^2] < \infty$.
If $E[\mid X \mid] = 1$, then

$$P[\mid X \mid \geq \lambda] \geq \frac{(1-\lambda)^2}{E[X^2]}$$

which provides a lower bound to complement Chebyshev's Inequality.

PROBLEM 1.9.2. Prove that if Z is a standard normal r.v. then, for all $t > 0$,

$$\frac{1}{\sqrt{2\pi}} \frac{t}{1+t^2} e^{-t^2/2} \leq P[Z \geq t] \leq \frac{1}{\sqrt{2\pi}} \frac{1}{t} e^{-t^2/2}$$

SOLUTION. The inequality

$$P[Z \geq t] \leq \frac{1}{\sqrt{2\pi}} \frac{1}{t} e^{-t^2/2}$$

can be proved using the following argument

$$\begin{aligned} P[Z \geq t] &= \int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \leq \frac{1}{\sqrt{2\pi}} \int_t^\infty \frac{z}{t} e^{-z^2/2} dz \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{t} \int_t^\infty z e^{-z^2/2} dz = \frac{1}{\sqrt{2\pi}} \frac{1}{t} e^{-t^2/2}. \end{aligned}$$

The second inequality is a little bit harder to prove. Essentially, we need to use the fact that for $z \geq t$ it is

$$\frac{1}{t^2} \int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \geq \int_t^\infty \frac{1}{z^2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz.$$

In fact, using integration by parts on the second integral, we find

$$\int_t^\infty \frac{1}{z^2}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz = \frac{1}{\sqrt{2\pi}}\left[-\frac{1}{z}e^{-z^2/2}\,|_t^\infty - \int_t^\infty e^{-z^2/2}dz\right]$$
$$= \frac{1}{\sqrt{2\pi}}\left[\frac{1}{t}e^{-t^2/2} - \int_t^\infty e^{-z^2/2}dz\right].$$

Rearranging the terms, we have that

$$\frac{1}{\sqrt{2\pi}}\left(1+\frac{1}{t^2}\right)\int_t^\infty e^{-z^2/2}dz \geq \frac{1}{\sqrt{2\pi}}\frac{1}{t}e^{-t^2/2}$$

and, finally,

$$\frac{1}{t^2}\int_t^\infty \frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz \geq \frac{1}{\sqrt{2\pi}}\frac{1}{t}\frac{t^2}{1+t^2}e^{-t^2/2} = \frac{1}{\sqrt{2\pi}}\frac{t}{t^2+1}e^{-t^2/2}$$

as we were supposed to prove.

PROBLEM 1.9.3. Let $S_n = \sum_{j=1}^n X_j$ where $\{X_j : j = 1,\ldots,n\}$ are independent r.v.'s such that

(1) $E[X_j] = 0$, $j = 1, 2, \ldots, n$.

(2) $E[X_j^2] = \sigma_j^2$, $j = 1, 2, \ldots, n$.

(3) $E[|\, X_j \,|^k] \leq (k!/2)\sigma_j^2 c^{k-2}$; $k > 2$, $j = 1,\ldots,n$, $0 < c < \infty$.

Prove that

$$P[S_n \geq x] \leq \exp\left\{\frac{-x^2}{2(s_n^2+cx)}\right\}$$

where $s_n^2 = \sum_{j=1}^n \sigma_j^2$.

SOLUTION. Several intermediate results are needed in order to prove the statement above.

Proposition 1. *$e^x \geq 1 - x$ for all $x \geq 0$.*

Proof. Trivial. □

Lemma (Bernstein's Inequality). *For any r.v. X for which an mgf exists and positive numbers a, t, it is*

$$P[X \geq a] \leq e^{-at}M_X(t)$$

where $M_X(t)$ is the mgf of the random variable X.

Proof. For any non-negative and non-decreasing function $g(x)$, $x \in \mathbf{R}$, it is possible to write

$$P[X \geq a] = P[g(X) \geq g(a)],$$

then, using Chebyshev's Inequality, we get

$$P[X \geq a] = P[g(X) \geq g(a)] \leq \frac{E[g(X)]}{g(a)}.$$

Let $g(x) \equiv e^{tx}$, $t > 0$; this definition clearly satisfies the requirements for $g(\cdot)$ given abovev. Thus,

$$P[X \geq a] = P[e^{tX} \geq e^{ta}] \leq e^{-ta} E[e^{tX}] = e^{-ta} M_X(t)$$

using the definition of mgf. □

Proposition 2. $E[e^{tX_j}] \leq \exp\{\sigma_j^2 t^2/(2(1-tc))\}$, $0 < tc < 1$, $j = 1, \ldots, n$.

Proof. Using a Taylor series expansion of e^{tx} at $x = 0$, one can convince oneself that for any $t > 0$ it must be

$$e^{tX_j} \leq 1 + tX_j + \sum_{k=2}^{\infty} \frac{t^k \mid X_j \mid^k}{k!}$$

Using the assumptions (1), (2) and (3) and taking the expectation on both sides of the inequality above we find that

$$E[e^{tX_j}] \leq 1 + t \cdot 0 + \sum_{k=2}^{\infty} \frac{t^k}{k!} \frac{k!}{2} \sigma_j^2 c^{k-2} = 1 + \frac{\sigma_j^2 t^2}{2} \left[\frac{1}{1-tc}\right]$$

as long as $0 < tc < 1$. This is always possible as it suffices taking $t < 1/c$ for any $c > 0$. Using the first Proposition it is now easily seen that

$$E[e^{tX_j}] \leq \exp\left\{\frac{\sigma_j^2 t^2}{2}\left[\frac{1}{1-tc}\right]\right\}$$

as we were supposed to show. □

Now, using a simple property of mgf's together with Proposition 2, we find that

$$\begin{aligned} M_{S_n}(t) = \Pi_{j=1}^n M_{X_j}(t) = \Pi_{j=1}^n E[e^{tX_j}] \leq \Pi_{j=1}^n \exp\left\{\frac{\sigma_j^2 t^2}{2}\left[\frac{1}{1-tc)}\right]\right\} \\ = \exp\left\{t^2 \frac{\sum_{j=1}^n \sigma_j^2}{2}\left[\frac{1}{1-tc}\right]\right\} = \exp\left\{\frac{s_n^2 t^2}{2}\left[\frac{1}{1-tc}\right]\right\}. \end{aligned}$$

Finally, using Bernstein's Inequality, we get

$$P[S_n \geq x] \leq \exp\{-tx\} \cdot \exp\{\frac{s_n^2 t^2}{2}\left(\frac{1}{1-tc}\right)\},$$

and, letting $t = x^2/(s_n^2 + cx)$, after simple algebraic manipulations we find

$$P[S_n \geq x] \leq \exp\{\frac{-x^2}{2(s_n^2 + cx)}\}$$

as we were asked to prove.

PROBLEM 1.9.4. Let $\{X_i : i = 1, \dots, n\}$ be i.i.d. r.v.'s such that $P[X_1 = 1] = 1 - P[X_1 = -1] = p$. Prove that $\forall \epsilon > 0$ there exists a positive number $K(\epsilon)$ such that

$$P\left[\left|\frac{S_n}{n} - (2p-1)\right| \geq \epsilon\right] \leq 2e^{-nK(\epsilon)}$$

where $S_n = X_1 + X_2 + \ldots + X_n$.

SOLUTION. Let a be any positive number, then

$$P[| S_n/n | \geq a] = P[S_n \geq na] + P[S_n \leq -na]$$

and, using Bernestein's Inequality, we have also:

$$P[S_n \geq na] \leq \inf_{t>0} e^{-nta} M_{S_n}(t)$$

$$P[S_n \leq -na] \leq \inf_{t>0} e^{nta} M_{S_n}(-t).$$

In this problem the mgf for the X_i's is easily computed as

$$M_{X_1}(t) = e^t p + e^{-t}(1-p),$$

therefore, simple properties of mgf's give

$$M_{S_n}(t) = [e^t p + e^{-t}(1-p)]^n.$$

Using this we can now write

$$\begin{aligned} P[S_n \geq na] &\leq \inf_{t>0} e^{-nta} M_{S_n}(t) = \inf_{t>0} e^{-nta}[e^t p + e^{-t}(1-p)]^n \\ &= \inf_{t>0} \exp\{-n[at - \log[e^t p + e^{-t}(1-p)]]\} \\ &= \exp\{-n \sup_{t>0}[at - \log[e^t p + e^{-t}(1-p)]]\}. \end{aligned}$$

The problem then reduces to maximizing the following function

$$f(t) = at - \log[e^t p + e^{-t}(1-p)].$$

It is a simple calculus routine to verify that the maximum for $f(\cdot)$ is achieved for

$$t' = \log\left[\frac{(1-p)(1+a)}{p(1-a)}\right]^{1/2}.$$

It is also easily checked that $t' > 0$ whenever $a > 2p-1$. Besides, one can verify that

$$f(t') = a\log\left[\frac{(1-p)(1+a)}{p(1-a)}\right]^{1/2} - \\ -\log\left[\left(\frac{p(1-p)(1-a)}{1+a}\right)^{1/2} + \left(\frac{p(1-p)(1+a)}{1-a}\right)^{1/2}\right].$$

Using the same procedure we can compute

$$\begin{aligned}\inf_{t>0} e^{-nat} M_{S_n}(-t) &= \inf_{t>0} e^{nat}[e^{-t}p + e^t(1-p)]^n \\ &= \inf_{t>0} \exp\{-n[-at - \log[e^{-t}p + e^t(1-p)]\} \\ &= \exp\{-n\sup_{t>0}(-at - \log[e^{-t}p + e^t(1-p)]\}.\end{aligned}$$

This is equivalent to maximize the function

$$g(t) = -at - \log[e^{-t}p + e^t(1-p)]\}$$

and, once again, a calculus routine gives that the maximum is achieved for

$$t'' = \log\left[\frac{p(1-a)}{(1-p)(1+a)}\right]^{1/2}$$

and $t'' > 0$ when $a < 2p-1$.

We have now all the elements to prove the statement of this problem. In fact,

$$P\left[\left|\frac{S_n}{n} - (2p-1)\right| \geq \epsilon\right] = P\left[\frac{S_n}{n} \geq 2p-1+\epsilon\right] + P\left[\frac{S_n}{n} \leq 2p-1-\epsilon\right].$$

Now, if we let $a = 2p-1+\epsilon$, we know from previous computations that

$$P\left[\frac{S_n}{n} \geq 2p-1+\epsilon\right] \leq \exp\{-nH_1(\epsilon)\}$$

where $H_1(\epsilon) = f(t')$ and $t' = \log[(1-p)p^{-1}(2p+\epsilon)(2-2p-\epsilon)^{-1}]^{1/2}$. Similarly, letting $a = 2p-1-\epsilon$, we find that

$$P\left[\frac{S_n}{n} \leq 2p-1-\epsilon\right] \leq \exp\{-nH_2(\epsilon)\}$$

where $H_2(\epsilon) = g(t'')$ and $t'' = \log[p(1-p)^{-1}(2-2p+\epsilon)(2p-\epsilon)^{-1}]^{1/2}$. Letting

$$K(\epsilon) = \min\{H_1(\epsilon), H_2(\epsilon)\}$$

it follows that

$$P\left[\left|\frac{S_n}{n} - (2p-1)\right| \geq \epsilon\right] \leq 2e^{-nK(\epsilon)}$$

as we were supposed to show. This clearly shows that $S_n/n \xrightarrow{P} E[X_1]$ at an exponential rate.

1.10 Moments and Orthogonal Transformations

PROBLEM 1.10.1. Consider the $(p+1) \times 1$ dimensional vector $(Y \quad X)$ where Y is a scalar random variable and X is a $p \times 1$ normal random vector with mean 0 and covariance I. Assume that there is a fixed $p \times k$, $(1 \leq k \leq p)$ matrix Q with orthogonal columns, $Q'Q = I$, such that the conditional density of $Y \mid X$, $f(y \mid X = x)$, is equal to the conditional density of Y given $Q'X$, $f(y \mid Q'X = Q'x)$; that is $f(y \mid X = x) = f(y \mid Q'X = Q'x)$ for all x.
Show that $E[X \mid Y = y]$ is in the subspace spanned by the columns of Q for all y.

SOLUTION. Let P be a $p \times (p-k)$ matrix with orthogonal columns and such that $P'Q = Q'P = 0$. This is always possible because of the Gram-Schmidt process. Letting $\Gamma = (P, Q)$, it is easily seen that $\Gamma'\Gamma = \Gamma\Gamma' = I$.
Using this fact we can write$E[X \mid Y]$ as

$$\begin{aligned} E[X \mid Y] &= E[\Gamma\Gamma' X \mid Y] = \Gamma E[\Gamma' X \mid Y] \\ &= PE[P'X \mid Y] + QE[Q'X \mid Y] \end{aligned}$$

so, if we can show that $E[P'X \mid Y] = 0$, we are done as in this case $E[X \mid Y] = QE[Q'X \mid Y] \in \mathcal{C}(Q)$, the subspace spanned by the columns of Q.
In order to do so we observe that as Γ is an orthogonal matrix and X is normally distributed with mean 0 and variance I_p, it is possible to write

$$f(x)dx = f(\Gamma' x)d(\Gamma' x).$$

Using the fact that $P'X$ and $Q'X$ are independent, we can also write

$$f(x)dx = f(P'x)f(Q'x)d(P'x)d(Q'x).$$

This fact and the assumption that $f(y \mid Q'x) = f(y \mid x)$ can now be used as follows:

$$\begin{aligned} E[P'X \mid Y] &= \int_{\mathbf{R}} P'x \frac{f(y \mid x) f(x)}{f(y)} dx \\ &= \int_{\mathbf{R}} \int_{\mathbf{R}} P'x f(y \mid Q'x) f(P'x) f(Q'x) d(P'x) d(Q'x) \\ &= \int_{\mathbf{R}} \int_{\mathbf{R}} P'x f(Q'x \mid y) f(P'x) f(Q'x) d(P'x) d(Q'x) \\ &= \int_{\mathbf{R}} f(Q'x \mid Y) d(Q'x) \int_{\mathbf{R}} P'x f(P'x) d(P'x) = 0, \end{aligned}$$

as $P'x \sim N_{p-k}(0, I)$, which gives the result we were looking for.

PROBLEM 1.10.2. Let X be a random variable in $\mathbf{R}^n$ and suppose that X and ΓX have the same distribution for every orthogonal matrix Γ. Prove the following:

(i) if $\mu = E[X]$ exists, then it must be $\mu = 0$;

(ii) if $\Sigma = Cov[X]$ exists, then $\Sigma = cI_n$, where I_n is the $n \times n$ identity matrix and $c \geq 0$.

SOLUTION. We know that for every choice of the matrix A, we have

$$E[AX] = AE[X] = A\mu.$$

Then, since by assumption X and ΓX have the same distribution, it must be $E[X] = \Gamma E[X]$ for every choice of the orthogonal matrix Γ. In particular, $-I_n$ is an orthogonal matrix and, thus, we have $\mu = -\mu$ which is possible iff $\mu = 0$.

It is also known that $Cov[AX] = A\Sigma A'$ and this holds for every choice of the matrix A. Then, since X and ΓX have the same distribution, it must be

$$\Sigma = \Gamma\Sigma\Gamma'$$

for any $n \times n$ orthogonal matrix Γ. In particular, this must be true for the orthogonal matrix $\Gamma_\star$ such that

$$\Gamma_\star\Sigma\Gamma_\star' = D = diag(\lambda_1, \lambda_2, \ldots, \lambda_n),$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of Σ. This is a consequence of the fact that Σ is a symmetric matrix and, therefore, the Spectral Theorem applies. It follows from this fact that Σ must be a diagonal matrix, but this is not all. In fact, let $x_1, x_2, \ldots, x_n$ be the eigenvectors of Σ associated with the eigenvalues above. We can assume that these eigenvectors have modulus one or, otherwise, we can rescale them satisfy this requirement. Then, we find that

$$x_i'\Sigma x_i = \lambda_i \parallel x_i \parallel^2 = \lambda_i$$

and, similarly, that

$$x_k'\Sigma x_k = \lambda_k$$

for all $i, k = 1, 2, \ldots, n$ such that $i \neq k$. In addition, we know that $\Sigma = \Gamma\Sigma\Gamma'$ for all orthogonal matrix Γ and this fact, in turn, allows us to write

$$x_i'\Gamma\Sigma\Gamma' x_i = \lambda_i, \ \forall i = 1, 2, \ldots, n.$$

This last equality can also be rewritten in the form

$$(\Gamma' x_i)'\Sigma(\Gamma' x_i) = \lambda_i$$

and, therefore, if we can prove that there exists an orthogonal matrix, say Ψ, such that $\Psi x_i = x_k$ we are done, because in that case $\lambda_i = \lambda_k$ for all choices of i and k. This is proved in the following

Proposition. *For every $x, y \in \mathbf{R}^n$, it is possible to find an orthogonal matrix Γ such that $\Gamma x = y$.*

Proof. Since x is a vector of length one, we can find an orthogonal matrix T such that $T'x = \epsilon_1$, with $\epsilon_1' = (1, 0, \ldots, 0)$. To check that this is the case let x be the first column of T and then complete the matrix with other $n-1$ column vectors, so that T is an orthogonal matrix. That this is possible it is clear using the Gram-Schmidt Orthogonalization process. The same can be done for y. Thus, we have found two orthogonal matrices, T_1 and T_2, such that

$$T_1'x = \epsilon_1, \quad T_2'y = \epsilon_1.$$

It is easily seen that this implies that $y = T_2T_1'x$ and since the product of two orthogonal matrices is still an orthogonal matrix we are done. □

1.11 Moments of Random Functionals, Stochastic Integrals, Markov chains

PROBLEM 1.11.1. For $0 \leq t \leq 1$, let $Z_t(\omega)$ be a random variable defined on a probability space $(\Omega, \mathcal{F}, P)$. Assume that $Z_t \sim N(\mu t, \sigma^2 t)$ for some constant μ and σ^2 and also assume that $Z_t(\omega)$ is a continuous function of t for every $\omega \in \Omega$. Define:

$$I(\omega) = \int_0^1 Z_t(\omega)dt \quad \omega \in \Omega.$$

Then,

1. $Z_t(\omega)$ is a measurable function of (t, ω) on the product space $([0,1] \times \Omega, \mathcal{B} \otimes \mathcal{F}, \lambda \otimes P)$, where $\mathcal{B}$ is the Borel σ-field on $[0,1]$, λ is Lebesgue measure on $[0,1]$ and P is the probability measure associated with the assumption of a normal distribution for $Z_t(\omega)$.
2. $I(\omega)$ is a random variable and $E[I] = \mu/2$.

SOLUTION. If we let

$$Z_{t,n}(\omega) = \sum_{k=1}^{n} Z_{k/n}(\omega) I_{[k/n,(k+1)/n]}(t),$$

we see that $Z_{t,n}(\omega)$ is a sum of products of functions which are $\mathcal{B} \otimes \mathcal{F}$-measurable. It follows that $Z_{t,n}(\omega)$ is measurable in (t, w), $n = 1, 2, \ldots$ and since $Z_t(\omega)$ can be written as a limit of measurable functions:

$$Z_t(\omega) = \lim_{n \to \infty} Z_{t,n}(\omega),$$

it is measurable as well.
In addition, $Z_t(\omega)$ is integrable with respect to $\lambda \otimes P$ as by Tonelli's Theorem we have that

$$\int_{[0,1]\times\Omega} \mid Z_t(\omega) \mid dP(\omega)dt = \int_{[0,1]} \left(\int_{\Omega} \mid Z_t(\omega) \mid dP(\omega) \right) dt = \int_0^1 E[\mid Z_t \mid]dt$$

$$\leq \left(\int_0^1 E[Z_t^2]dt\right)^{1/2} = \left(\int_0^1 (\mu^2t^2+\sigma^2t)dt\right)^{1/2} < \infty.$$

Since λ and P are clearly σ-finite measures, Fubini's Theorem applies and we can conclude that:

- $I(\omega) = \int_0^1 Z_t(\omega)dt$ is $\mathcal{F}$-measurable and hence a random variable;
- $E[I] = \int_0^1 E[Z_t]dt = \int_0^1 \mu t dt = \mu/2$

as we were supposed to show.

PROBLEM 1.11.2. Let $W(t)$, $0 \leq t < \infty$, be the Wiener Process with parameter σ^2.

1. Find the mean and variance of $\int_0^1 W^2(t)dt$.
2. Prove that the correlation between $W(t)$ and $\int_0^1 W(s)ds$ is $\sqrt{3t}(2-t)/2$.
3. Set $X(t) = \int_0^t W(s)ds$, $t \geq 0$. Find the mean and covariance function of the process $X(t)$.
4. Let $X = \int_0^1 tdW(t)$ and $Y = \int_0^1 t^2dW(t)$. Find the mean variance of X and Y. Prove also that the correlation between these two random variables is $\sqrt{15}/4$.

SOLUTION. To compute the mean for $\int_0^1 W^2(t)dt$ is an easy task: in fact,

$$E[\int_0^1 W^2(t)dt] = \int_0^1 E[W^2(t)]dt = \int_0^1 \sigma^2 tdt = \sigma^2/2$$

since the existence of the integral above allows interchanging the operators E and $\int$. To compute the second moment, we need to use the fact that for the Wiener process the increments $W(t_2) - W(t_1), \ldots, W(t_n) - W(t_{n-1})$ are independent for $t_1 \leq t_2 \leq \ldots \leq t_{n-1} \leq t_n$. Then,

$$\begin{aligned} E[\int_0^1 W^2(t)dt]^2 &= E[\int_0^1 W^2(t)dt \int_0^1 W^2(s)ds] = \\ &= E[\int_0^1 (\int_0^1 W^2(t)W^2(s)dt)ds] \end{aligned}$$

and, after interchanging of E and $\int$,

$$\begin{aligned} &= \int_0^1 (\int_0^1 E[W^2(t)W^2(s)]dt)ds \\ &= \int_0^1 (\int_0^s E[W^2(t)W^2(s)]dt)ds + \int_0^1 (\int_s^1 E[W^2(t)W^2(s)]dt)ds. \end{aligned}$$

When $t \leq s$, one can write $W(s) = [W(s) - W(t) + W(t)]$ so that the first of the last two integrals above can be rewritten as

$$\int_0^1 (\int_0^s E[W^2(t)W^2(s)]dt)ds = \int_0^1 (\int_0^s E[W^2(t)\{W(s) - W(t) + W(t)\}]^2 dt)ds$$

$$= \int_0^1 (\int_0^s E[W^2(t)\{W(s) - W(t)\}^2 + 2\{W(s) - W(t)\}W^3(t) + W^4(t)]dt)ds$$

and, using independence of $W(s) - W(t)$ and $W(t)$ together with basic properties of the Wiener process, we are led to write

$$= \sigma^4 \int_0^1 (\int_0^s [(s-t)t + 0 + 3t^2]dt)ds = \frac{7}{24}\sigma^4.$$

It is easily checked that the second of the two integrals above equals $(7/24)\sigma^4$ as well. In fact, since all integrals here exist, by Fubini's Theorem we can write

$$\int_0^1 (\int_s^1 E[W^2(t)W^2(s)]dt)ds = \int_0^1 (\int_0^t E[W^2(s)W^2(t)]ds)dt.$$

This implies that

$$E[\int_0^1 W^2(t)dt]^2 = 2\sigma^4 \frac{7}{24} = \frac{7}{12}\sigma^4$$

and, therefore,

$$Var[\int_0^1 W^2(t)dt] = \frac{7}{12}\sigma^4 - \left(\frac{\sigma^2}{2}\right)^2 = \frac{\sigma^4}{3}.$$

To prove the second statement we need the following:

Lemma. *Under the usual assumptions (f and g continuously differentiable functions on suitable closed intervals),*

$$E[\int_a^b f(t)dW(t) \int_c^d g(t)dW(t)] = 0 \quad a \leq b \leq c \leq d,$$

and

$$E[\int_a^b f(t)dW(t) \int_a^c g(t)dW(t)] = \sigma^2 \int_a^b f(t)g(t)dt \quad a \leq b \leq c.$$

Proof. See Hoel, Port and Stone, 1972; pp. 144-5. □

As $W(t) \sim N(0, \sigma^2 t)$ and[11] $\int_0^1 W(s)ds \sim N(0, \sigma^2/3)$, one clearly finds that

$$E[W(t)] = E[\int_0^1 W(s)ds] = 0; \; Var[W(t)] = \sigma^2 t, \; Var[\int_0^1 W(s)ds] = \sigma^2/3.$$

[11]This follows from the second part of Problem 2.6.2 in Chapter 2. There it is proved that $\int g(t)W(t)dt \sim N(0, \sigma^2 B)$ where $B = \int_0^1 G^2(t) + G^2(1) - 2G(1)\int_0^1 G(t)dt$ and $G(t) = \int_0^t g(s)ds$, $0 \leq t \leq 1$. If we let $g(t) = 1$, $\forall t \in [0,1]$ then, $G(t) = t$ and $B = 1/3$.

The covariance between $W(t)$ and $\int_0^1 W(s)ds$ is given by $E[W(t) \cdot \int_0^1 W(s)ds]$. To compute this expectation we can use the formula for integration by parts proved in Problem 2.6.3 in Chapter 2 and write

$$W(t) = \int_0^t dW(z) \quad \text{and} \quad \int_0^1 W(s)ds = W(1) - \int_0^1 zdW(z),$$

Thus,

$$Cov[W(t), \int_0^1 W(s)ds] = E[W(1) \int_0^t dW(z)] - E[\int_0^t dW(z) \int_0^1 zdW(z)]$$

and using the lemma above, one has that

$$E[\int_0^t dW(z) \int_0^1 zdW(z)] = E[\int_0^t dW(z) \int_0^t zdW(z)]$$

$$= \sigma^2 \int_0^t zdz = \sigma^2 t^2/2.$$

Similarly,

$$E[W(1) \int_0^t dW(z)] = E[\int_0^1 dW(z) \int_0^t dW(z)] = \sigma^2 \int_0^t dz = \sigma^2 t.$$

Putting all this partial results together, we finally find

$$Corr(W(t), \int_0^1 W(s)ds) = \frac{-\sigma^2 t^2/2 + \sigma^2 t}{\sqrt{\sigma^2 t \times \sigma^2/3}} = \frac{\sqrt{3t}(2-t)}{2}$$

as we were supposed to show.

Assume first that $s \leq t$. It is easily checked that $E[X(t)] = 0$, $\forall t \geq 0$. Then, the covariance function between $X(t)$ and $X(s)$ is given by

$$E[X(t)X(s)] = \int_0^t (\int_0^s E[W(v)W(u)dv)du.$$

Now, depending on whether $v > u$ or $v < u$, we have

$$\int_0^t (\int_0^s E[W(v)W(u)dv)du = \int_0^s (\int_u^s \sigma^2 udv)du + \int_0^t (\int_0^{u \wedge s} \sigma^2 vdv)du.$$

Then,

$$\int_0^s (\int_u^s \sigma^2 udv)du = \int_0^s (\sigma^2 us - \sigma^2 u^2)du = \frac{\sigma^2 s^3}{6};$$

$$\int_0^t (\int_0^{u \wedge s} \sigma^2 vdv)du = \int_0^s (\int_0^u \sigma^2 vdv)du + \int_s^t (\int_0^s \sigma^2 vdv)du = \frac{\sigma^2 s^2 t}{2} - \frac{\sigma^2 s^3}{2}.$$

Using these last two results we have found that

$$Cov[X(s), X(t)] = \frac{\sigma^2 s^3}{6} + \frac{\sigma^2 s^2 t}{2} - \frac{\sigma^2 s^3}{2} = \frac{\sigma^2 s^2(3t - s)}{6}.$$

Instead, if $s > t$, the same argument gives:

$$Cov[X(s), X(t)] = \frac{\sigma^2 t^2(3s - t)}{6}.$$

For the last part of this problem we can use the result from Problem 2.16 which gives

$$E[X] = E[\int_0^1 tdW(t)] = E[1W(1) - \int_0^1 W(t)dt] = 0;$$

$$E[Y] = E[\int_0^1 t^2 dW(t)] = E[1W(1) - 2\int_0^1 W(t)dt] = 0.$$

To compute $E[X^2]$ we can use the fact also stated in Problem 2.16

$$E[X^2] = E[\int_0^1 tdW(t) \int_0^1 sdW(s)] = \int_0^1 t^2 dt = \frac{\sigma^2}{3}.$$

The same fact applied to $E[Y^2]$ gives $E[Y^2] = \sigma^2/5$. Finally,

$$Cov[X, Y] = E[\int_0^1 tdW(t) \int_0^1 t^2 dW(t)] = \sigma^2 \int_0^1 t^3 dt = \frac{\sigma^2}{4}.$$

Then, the correlation is easily computed as

$$Corr(X, Y) = \frac{\sigma^2/4}{(\sigma^2/2 \times \sigma^2/3)^{1/2}} = \frac{\sqrt{15}}{4}.$$

PROBLEM 1.11.3. Let $\{W_t : t \in \mathbf{R}_+\}$ be Browian motion such that $W_0 = 0$. Define

$$\beta_{2n}(t) = E[W_t^{2n}], \quad n = 0, 1, 2, \ldots; t \geq 0.$$

Prove that

$$\beta_{2n}(t) = n \cdot (2n - 1) \cdot \int_0^t \beta_{2n-2}(s)ds.$$

SOLUTION. There are several ways to prove this result. The most natural one is to compute the moments for $n = 1, 2, \ldots, k$ and use induction. Nevertheless, a much faster way to derive the result stated above uses basic stochastic calculus. In fact, using Itô's Formula and the fact that by assumption $W_0 = 0$, one finds that

$$W_t^{2n} = n \cdot (2n - 1) \int_0^t W_t^{2n-2} dt + 2n \int_0^t W_t^{2n-1} dW_t$$

and, since $W_t^{2n-2} \geq 0$, we can use Fubini's Theorem and interchange integration and expectation in the first integral on the right-hand side. This gives that

$$\beta_{2n}(t) \equiv E[W_t^{2n}] = n \cdot (2n-1) \int_0^t E[W_t^{2n-2}]dt + 2n \cdot E\Big[\int_0^t W_t^{2n-1} dW_t\Big].$$

Approximating the stochastic integral on the right-hand side by sums and proving that the sequence of sums converges in mean square to 0 completes the proof.

PROBLEM 1.11.4. Compute mean and variance for X_t when the evolution of X_t is described by the following stochastic differential equation

$$dX_t = \alpha(\beta - \gamma X_t)\, dt + \sigma X_t^{1/2}\, dW_t$$

and where $\alpha, \beta, \gamma, \sigma$ and X_0 are positive constants.

SOLUTION. To compute the mean one uses the fact that the stochastic differential equation above can be written in the integral form

$$X_t = X_0 + \alpha \int_0^t (\beta - \gamma X_u)\, du + \sigma \int_0^t X_u^{1/2}\, dW_u$$

from which, taking expectations and remembering that the expectation of Itô integrals is zero, we obtain

$$E[X_t] = X_0 + \alpha \int_0^t (\beta - \gamma E[X_u])\, du.$$

Differentiation with respect to t yields

$$\frac{d}{dt} E[X_t] = \alpha(\beta - \gamma E[X_t])$$

and this is easily checked to imply that

$$\frac{d}{dt}(e^{\alpha\gamma t} E[X_t]) = e^{\alpha\gamma t}\left[\alpha\gamma E[X_t] + \frac{d}{dt} E[X_t]\right] = \alpha\beta e^{\alpha\gamma t}.$$

Integration with respect to t yields

$$e^{\alpha\gamma t} E[X_t] - X_0 = \alpha\beta \int_0^t e^{\alpha\gamma u}\, du = \frac{\beta}{\gamma}\Big(e^{\alpha\gamma t} - 1\Big)$$

which can be solved for $E[X_t]$ and yields

$$E[X_t] = \frac{\beta}{\gamma} + E^{-\alpha\gamma t}\Big(X_0 - \frac{\beta}{\gamma}\Big).$$

To compute the variance we need to compute $E[X_t^2]$ first. To this purpose, Itô's formula can be used to yield

$$dX_t^2 = (2\alpha\beta + \sigma^2 - 2\alpha\gamma X_t^2)\,dt + 2\sigma X_t^{3/2}\,dW_t$$

which can be rewritten in integral form as

$$dX_t^2 = X_0^2 + (2\alpha\gamma + \sigma^2)\int_0^t X_u\,du - 2\alpha\gamma\int_0^t X_u^2\,du + 2\sigma\int_0^t X_u^{3/2}\,dW_u.$$

Taking expectations and recalling again that the expected value of Itô integrals is zero, we find

$$E[X_t^2] = X_0^2 + (2\alpha\beta + \sigma^2)\int_0^t E[X_u]\,du - 2\alpha\gamma\int_0^t E[X_u^2]\,du.$$

Differentiating with respect to t yields

$$\frac{d}{dt}E[X_t^2] = (2\alpha\beta + \sigma^2)E[X_t] - 2\alpha\gamma E[X_t^2],$$

and this, in turn, gives

$$\frac{d}{dt}e^{2\alpha\gamma t}E[X_t^2] = e^{2\alpha\gamma t}\left(2\alpha\gamma E[X_t^2] + \frac{d}{dt}E[X_t^2]\right) = e^{2\alpha\gamma t}(2\alpha\beta + \sigma^2)E[X_t].$$

Finally, using the formula derived above for $E[X_t]$ and integrating the last equation with respect to t one can show that

$$\begin{aligned} E[X_t^2] = {} & \frac{\beta\sigma^2}{2\alpha\gamma^2} + \frac{\beta^2}{\gamma^2} + \left(X_0 - \frac{\beta}{\gamma}\right)\left(\frac{\sigma^2}{\alpha\gamma} + \frac{2\beta}{\gamma}\right)e^{-\alpha\gamma t} \\ & + \left[\left(X_0 - \frac{\beta}{\gamma}\right)^2\frac{\sigma^2}{\alpha\gamma} + \frac{\sigma^2}{\alpha\gamma}\left(\frac{\beta}{2\gamma} - X_0\right)\right]e^{-2\alpha\gamma t}. \end{aligned}$$

The variance of X_t is then given by

$$\begin{aligned} \mathrm{Var}[X_t] &= E[X_t^2] - (E[X_t])^2 \\ &= \frac{\beta\sigma^2}{2\alpha\gamma^2} + \left(X_0 - \frac{\beta}{\gamma}\right)\frac{\sigma^2}{\alpha\gamma}e^{-\alpha\gamma t} + \frac{\sigma^2}{\alpha\gamma}\left(\frac{\beta}{2\gamma} - X_0\right)e^{-2\alpha\gamma t}. \end{aligned}$$

PROBLEM 1.11.5. Let $\{X_n : n \in \mathbf{Z}_+\}$ be Markov chain on the state space $\mathcal{S} = \{s_1, s_2, \ldots, s_N\}$ with initial distribution λ and transition matrix P. Find $E[X_i]$ and $\mathrm{Cov}[X_i, X_{i+j}]$ as a function of P and λ. Assume that

$$P = Q + \theta A,\quad A = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix},\quad Q = \begin{pmatrix} \alpha & 1-\alpha \\ \alpha & 1-\alpha \end{pmatrix}$$

and θ any real constant in $(0,1)$. Find an explicit expression for $\mathrm{Cov}(X_i, X_{i+j})$.

SOLUTION. By definition,

$$E[X_i] = \sum_{j=1}^{N} p_j \cdot s_j$$

where $p_j = Pr[X_i = s_j]$, $j = 1, 2, \ldots, N$. In addition, for a Markov chain, we have

$$Pr[X_i = s_j] = (\lambda^T P^i)_j$$

where P^i is the i-th power of the transition matrix P and $(\cdot)_j$ denotes the i-th element of the row vector $\lambda^T P^i$. This can be verified easily using the following argument:

$$\begin{aligned} Pr[X_i = s_j] &= \sum_{h_0 \in \mathcal{S}} \cdot \sum_{h_{i-1} \in \mathcal{S}} Pr[X_0 = h_0, \ldots, X_{i-1} = h_{i-1}, X_i = S_j] \\ &= \sum_{h_0 \in \mathcal{S}} \cdot \sum_{h_{i-1} \in \mathcal{S}} \lambda_{h_0} \cdot p_{h_0,h_1} \cdot \ldots \cdot p_{h_{i-1},s_j} = (\lambda^T P^i)_j. \end{aligned}$$

Thus, we have found that

$$E[X_i] = \sum_{j=1}^{N} p_j \cdot s_j = \lambda^T P^i s$$

with $s^T = (s_1, s_2, \ldots, s_N)$.

To establish the result for the covariance of two elements of the chain, X_i and X_{i+j}, we first compute $E[X_i X_{i+j}]$. Clearly, we have

$$E[X_i X_{i+j}] = \sum_{a \in \mathcal{S}} \sum_{b \in \mathcal{S}} ab \cdot p_{a,b}$$

where $p_{a,b} = Pr[X_i = a, X_j = b]$ and, using the Markov property,

$$p_{a,b} = (P^j)_{a,b} (\lambda^T P^i)_a$$

where $(A)_{a,b}$ denotes the (a, b) element of the matrix A. Now, if we rewrite $E[X_i X_{i+j}]$ in the form

$$E[X_i X_{i+j}] = \sum_{c \in \mathcal{S}} \sum_{a \in \mathcal{S}} a \cdot \lambda_a p^i_{a,c} \sum_{b \in \mathcal{S}} b \cdot p^j_{c,b}$$

this last expression can be written also as

$$\begin{aligned} E[X_i X_{i+j}] = &\left(\sum_{a \in \mathcal{S}} s_1 \cdot \lambda_a p^i_{a,1}, \sum_{a \in \mathcal{S}} s_2 \cdot \lambda_a p^i_{a,2}, \ldots, \sum_{a \in \mathcal{S}} s_N \cdot \lambda_a p^i_{a,N} \right) \\ &\times \left(\sum_{b \in \mathcal{S}} p^j_{1,b} s_b, \sum_{b \in \mathcal{S}} p^j_{2,b} s_b \ldots, \sum_{b \in \mathcal{S}} p^j_{N,b} s_b \right)^T. \end{aligned}$$

Now, it is not difficult to verify that

$$\left(\sum_{a\in S} s_1\cdot\lambda_a p^i_{a,1}, \sum_{a\in S} s_2\cdot\lambda_a p^i_{a,2}, \ldots, \sum_{a\in S} s_N\cdot\lambda_a p^i_{a,N}\right) = (\lambda^T P^i)\mathrm{diag}(s_1, s_2, \ldots, s_N)$$

while

$$\left(\sum_{b\in S} p^j_{1,b}s_b, \sum_{b\in S} p^j_{2,b}s_b \ldots, \sum_{b\in S} p^j_{N,b}s_b\right)^T = P^j s.$$

Putting these two last facts together makes it possible to write

$$E[X_i X_{i+j}] = \lambda^T P^i \mathrm{diag}(s_1, s_2, \ldots, s_N) P^j s.$$

Now, since $P^{i+j} = P^i P^j$, the expression for the covariance can be further simplified to look like

$$\begin{aligned}\mathrm{Cov}[X_i X_{i+j}] &= E[X_i X_{i+j}] - E[X_i]E[X_{i+j}]\\ &= \lambda^T P^i \mathrm{diag}(s_1, s_2, \ldots, s_N) P^j s - \lambda^T P^i s \lambda^T P^{i+j} s\\ &= \lambda^T P^i[\mathrm{diag}(s_1, s_2, \ldots, s_N) - s\lambda^T P^i] P^j s\end{aligned}$$

When $P = Q+\theta A$ and A, Q are defined above, one can easily verify the following facts

(1) $QA = \emptyset$, where $\emptyset$ is a 2×2 null matrix;

(2) $P^i = Q + 2^{i-1}\theta^i A$, $i = 1, 2, \ldots$;

(3) $[\mathrm{diag}(s_1, s_2) + s\lambda^T P^i]e = 0$, $i = 1, 2, \ldots$ where $e^T = (1, 1)$ and $0^T = (0, 0)$

(4) $Q = e(\alpha, 1-\alpha)$.

Now, let $\lambda^T P^i \equiv (\rho_i, 1-\rho_i)^T \equiv \rho^T$. Then

$$\begin{aligned}\mathrm{Cov}(X_i, X_{i+j}) &= \lambda^T P^i[\mathrm{diag}(s_1, s_2) - s\lambda^T P^i]P^j s = \rho^T[\mathrm{diag}(s_1, s_2) - s\rho^T]P^j s\\ &\overset{(2)}{=} \rho^T[\mathrm{diag}(s_1, s_2) - s\rho^T](Q + 2^{j-1}\theta^j A)s\\ &\overset{(4)}{=} \rho^T[\mathrm{diag}(s_1, s_2) - s\rho^T]e(\alpha, 1-\alpha)s + 2^{j-1}\theta^j\rho^T[\mathrm{diag}(s_1, s_2) - s\rho^T]As\\ &\overset{(3)}{=} 2^{j-1}\theta^j\rho^T[\mathrm{diag}(s_1, s_2) - s\rho^T]As = 2^j\theta^j\rho_i(1-\rho_i)(s_1 - s_2)^2.\end{aligned}$$

2 Chapter DISTRIBUTION FUNCTIONS AND TRANSFORMATIONS

2.1 Transformations of Discrete Random Variables

PROBLEM 2.1.1. Let X be an exponential(1) random variable, and define Y to be the integer part of X+1, that is

$$Y = i+1 \text{ iff } i \le X < i+1, \quad i = 1, 2, \ldots$$

(a) Find the distribution of Y. What well-known distribution does Y have ?

(b) Find the conditional distribution of X-4 given $Y \ge 5$.

SOLUTION. Since $X \sim$ exponential(1) we have that $f_X(x) = \mathrm{e}^{-x}$ for all $x \ge 0$. By definition, we have that $Y = i+1$ iff $i \le X < i+1$ where i ranges from 1 to ∞. This means that the possible values for Y are $\{1, 2, \ldots, \}$. Therefore, we find that

$$P(Y=y) = P(y-1 \le X < y+1) = \int_{y-1}^{y} \mathrm{e}^{-x} dx = \mathrm{e}^{-y}(\mathrm{e}-1) = \left(1 - \frac{1}{\mathrm{e}}\right) \cdot \left(\frac{1}{\mathrm{e}}\right)^{y-1}.$$

This tells us that

$$f_Y(y) = \left(1 - \frac{1}{\mathrm{e}}\right) \cdot \left(\frac{1}{\mathrm{e}}\right)^{y-1} I_{\{1,2,\ldots\}},$$

and thus we are led to conclude that Y is distributed as a geometric(1 − 1/e) random variable.

Let now $Z = X - 4$. Then, we get

$$F_{Z|Y\ge 5}(z) = \frac{P(Z \le z; Y \ge 5)}{P(Y \ge 5)} = \frac{P(X \le z+4; X \ge 4)}{P(Y \ge 5)} = \frac{P(4 \le X \le z+4)}{P(X \ge 4)}.$$

This last expression is very easy to deal with as X is an exponential(1) random variable. This gives

$$F_{Z|Y\geq 5}(z) = \frac{e^{-4} - e^{-(4+z)}}{e^{-4}} = (1 - e^{-z})I_{[0,\infty)}(z).$$

(It should be evident from the structure of the random variable Z that if $z \in [-4, 0)$, then $F_{Z|Y\geq 5}(z) = 0$.)
To find the p.d.f. of this random variable it suffices to take the first derivative of the distribution above. This gives:

$$f_{Z|Y\geq 5}(z) = \frac{dF_{Z|Y\geq 5}(z)}{dz} = \begin{cases} e^{-z} & \text{if } z \geq 0; \\ 0 & \text{if } -4 \leq z < 0. \end{cases}$$

This completes the problem.

PROBLEM 2.1.2. Let X and Y be independent random variables with the same geometric distribution.

(a) Show that U and V are independent, where U and V are defined by

$$U = \min\{X, Y\}, \quad \textit{and} \quad V = X - Y.$$

(b) Find the distribution of Z = X/(X+Y), where we define Z = 0 if X + Y = 0.

(c) Find the joint mgf of X and $X + Y$.

SOLUTION. For part (a) we use the fact that for any two numbers it is:

$$\min\{X, Y\} = \frac{X + Y - \mid X - Y \mid}{2}.$$

Then, we find that:

$$f_{U,V}(u, v) = P(U = u, V = v) = P\Big(\frac{X + Y - \mid X - Y \mid}{2} = u, X - Y = v\Big)$$

and therefore

$$P(U = u, V = v) = P(X + Y = 2u + \mid v \mid, X - Y = v)$$

where $u \in \{1, 2, \ldots,\}$ and $v \in \{\ldots, -n, -n + 1, \ldots, n - 1, n, \ldots\}$.
We need now to compute the joint pmf for $X + Y$ and $X - Y$. This is done as follows:

$$P(X + Y = z, X - Y = w) = P(X = (z + w)/2, Y = (z - w)/2)$$

and using the assumption of independence between X and Y we get

$$P(X + Y = z, X - Y = w) = P(X = (z + w)/2) \cdot P(Y = (z - w)/2).$$

We know that both X and Y are identically distributed as geometric(p) and, therefore, it is possible to write

$$P(X+Y=z,X-Y=w)=p(1-p)^{(z+w)/2-1}\cdot p(1-p)^{(z-w)/2-1}=p^2(1-p)^{z-2}.$$

Using this result we have

$$\begin{aligned}P(U=u,V=v)&=P(X+Y=2u+|v|,X-Y=v)\\&=p^2(1-p)^{2u-2}(1-p)^{|v|}\times I_{\{1,2,\ldots\}}(u)I_{\{\ldots,-n,-n+1,\ldots,n-1,n,\ldots\}}(v).\end{aligned}$$

Therefore, as the pmf for U and V is such that

$$f_{U,V}(u,v)=g(u)h(v),$$

with

$$g(u)=p^2(1-p)^{2u-2}I_{\{1,2,\ldots\}}(u)$$

and

$$h(v)=(1-p)^{|v|}I_{\{\ldots,-n,-n+1,\ldots,n-1,n,\ldots\}}(v),$$

we are allowed to conclude that U and V are independent.
This result is also useful in establishing another fact, i.e.:

$$\min\{X,Y\}\sim\ \text{geometric}(1-(1-p)^2).$$

In fact, we have that

$$\begin{aligned}f_U(u)&=p^2(1-p)^{2u-2}I_{\{1,2,\ldots\}}(u)\sum_{v=-\infty}^{\infty}(1-p)^{|v|}\\&=p^2(1-p)^{2u-2}I_{\{1,2,\ldots\}}(u)\cdot\frac{(2-p)}{p}\\&=(2p-p^2)(1-p)^{2u-2}I_{\{1,2,\ldots\}}(u)\\&=(1-(1-p)^2)[(1-p)^2]^{u-1}I_{\{1,2,\ldots\}}(u)\end{aligned}$$

which is the pmf of a geometric$(1-(1-p)^2)$ random variable.

Part (b) is similar to part (a). In fact, we have that

$$\begin{aligned}f_{Z,X}(z,x)&=P(Z=z,X=x)=P\Big(\frac{X}{X+Y}=z,X=x\Big)\\&=P\Big(Y=x/z-x,X=x)=P(X=x)\cdot P(Y=x/z-x)\end{aligned}$$

because of independence of X and Y.
Finally we get

$$f_{Z,X}(z,x)=p(1-p)^{x-1}\cdot p(1-p)^{x/z-x-1}I_{(0,1)\cap\mathbf{Q}}(z)\cdot I_{\{1,2,\ldots\}}(x).$$

As we are interested only in the marginal distribution for Z, we have to sum out X. Thus,

$$
\begin{aligned}
f_Z(z) &= \sum_{x=1}^{\infty} \left(\frac{p}{1-p}\right)^2 (1-p)^{x/z} \\
&= \left(\frac{p}{1-p}\right)^2 \frac{(1-p)^{1/z}}{1-(1-p)^{1/z}} I_{(0,1)\cap\mathbf{Q}}(z).
\end{aligned}
$$

Part (c) is straigthforward as we have

$$
\begin{aligned}
M_{X,X+Y}(t_1,t_2) &= E[\exp\{t_1X + t_2(X+Y)\}] = E[\exp\{(t_1+t_2)X + t_2Y\}] \\
&= E[\exp\{(t_1+t_2)X] + E[\exp\{t_2Y\}] = M_X(t_1+t_2)\cdot M_Y(t_2).
\end{aligned}
$$

Finally, since for a geometric(p) random variable the mgf is given by

$$
M_X(t) = \frac{pe^t}{1-(1-p)e^t} \quad t < -\log(1-p),
$$

we have that

$$
\begin{aligned}
M_{X,X+Y}(t_1,t_2) &= \frac{pe^{t_1+t_2}}{1-(1-p)e^{t_1+t_2}} \cdot \frac{e^{t_2}}{1-(1-p)e^{t_2}} \\
&= \frac{pe^{t_1+2t_2}}{[1-(1-p)e^{t_1+t_2}][1-(1-p)e^{t_2}]}.
\end{aligned}
$$

PROBLEM 2.1.3. Suppose we observe counts $\{n_i, i = 1,2,\ldots,N\}$ in the N cells of a contingency table. For istance, these might be observations for the N levels of a single categorical variable, or for N = IJ cells of a two-way table. We treat the counts as random variables. Each n_i has distribution concentrated on the nonnegative integers.

The Poisson sampling model for counts $\{n_i\}$ assumes that they are independent Poisson random variables.

Let X_i, $i = 1,2,\ldots,N$, be N-random variables whose pmf is given by poisson(λ_i) and let $X = \sum_i X_i$ be the total sample size.

Prove the following facts:

a. $X_i \mid X \sim \text{binomial}(x, \lambda_i/(\sum_i^N \lambda_i))$;

b. $X_i, X_j \mid X \sim \text{multinomial}(x, \lambda_i/(\sum_i^N \lambda_i), \lambda_j/(\sum_i^N \lambda_i))$;

c. $X_1, X_2, \ldots X_n \mid X \sim \text{multinomial}(x, p_1, p_2, \ldots, p_n)$, $n = 1,2,\ldots,N$ with $p_i = \lambda_i/(\sum_i^N \lambda_i)$, $i = 1,2,\ldots,N$.

SOLUTION. As $X_i \sim$ poisson(λ_i), $I = 1, 2, \ldots, N$, where the X_i's are pairwise independent r.v. we know that $X = \sum_i X_i \sim$ poisson$(\sum_i \lambda_i)$. The proof of this fact can be provided using c.f.s. In fact, $\phi_X(t) = \phi_{\sum_i^n X_i}(t) = \Pi_{i=1}^n \phi_{X_i}(t)$ and, using the analytical form of the c.f. for a Poisson r.v., we find

$$\phi_{\sum_i^n X_i}(t) = \Pi_{i=1}^n e^{\lambda_i(e^{it}-1)} = e^{\sum_{i=1}^n \lambda_i(e^{it}-1)}.$$

Similarly, $X - X_i \sim$ poisson$(\sum_{j=1, j\neq i}^N \lambda_j)$. Thus, we have:

$$f_{X_i|X}(x_i) = P(X_i = x_i \mid X = n) = \frac{P(X_i = x_i, \sum_{i=1}^N = x)}{P(\sum_{i=1}^N = x)}$$
$$= \frac{P(X_i = x_i, \sum_{j=1, j\neq i}^N X_j = x - x_i)}{P(\sum_{j=1}^N X_i = x)}.$$

Independence of $\sum_{j=1, j\neq i}^N X_i$ and X_i (see e.g., Casella and Berger, Theorem 4.3.2) together with some algebra suffice to show that

$$f_{X_i|X}(x_i) = \binom{x}{x_i}\left(\frac{\lambda_i}{\sum_i^N \lambda_i}\right)^{x_i}\left(1 - \frac{\lambda_i}{\sum_i^N \lambda_i}\right)^{x-x_i}.$$

Thus, it is now possible to state that

$$X_i \mid X \sim \text{binomial}\left(x, \frac{\lambda_i}{\sum_i^N \lambda_i}\right)$$

A similar technique works for part (b) as well. In fact, we have

$$f_{X_i=x_i, X_j=x_j|X=x}(x_i, x_j) = P(X_i = x_i, X_j = x_j, \sum_i^N X_i = x) =$$
$$= \frac{P(X_i = x_i, X_j = x_j, \sum_{h=1, h\neq i,j}^N = x - x_i - x_j)}{P(\sum_{i=1}^N X_i = x)}.$$

Using again Theorem 4.3.2 we are led to the following result:

$$f_{X_i=x_i, X_j=x_j|X=x}(x_i, x_j) =$$
$$\frac{x!}{x_i! x_j! (x - x_i - x_j)!}\left(\frac{\lambda_i}{\sum_{i=1}^N \lambda_i}\right)^{x_i} \cdot \left(\frac{\lambda_j}{\sum_{i=1}^N \lambda_i}\right)^{x_j} \cdot \left(1 - \frac{\lambda_i + \lambda_j}{\sum_{i=1}^N \lambda_i}\right)^{x-x_i-x_j}.$$

Therefore we have found that

$$X_i, X_j \mid X \sim \text{multinomial}\left(x, \frac{\lambda_i}{\sum_{i=1}^N \lambda_i}, \frac{\lambda_j}{\sum_{i=1}^N \lambda_i}\right).$$

The general result is now easily established: we can prove by finite induction that

$$X_1, X_2, \ldots X_n \mid X \sim \text{multinomial}(x, p_1, p_2, \ldots, p_n), \quad n = 1, 2, \ldots, N$$

with $p_i = \lambda_i / (\sum_i^N \lambda_i)$, $i = 1, 2, \ldots, N$.

2.2 Transformations of Exponential Random Variables

PROBLEM 2.2.1. Prove that if $X_i \sim$ exponential(λ) $i = 1, 2, \dots, n$, and X_i is independent of X_j for every $i \neq j$, then

(a) $\sum_{i=1}^n X_i \sim$ gamma(n, λ);

(b) $X_i - X_j \sim$ double exponential$(0, \lambda)$;

(c) $\min_i\{X_i\} \sim$ exponential(λ/n).

SOLUTION. The easiest way to prove (a) is to use a moment generating function technique. In fact, we know that

$$M_{X_i}(t) = \frac{1}{1 - \lambda t} \;\; \forall t < \frac{1}{\lambda} \;\; \forall i = 1, 2, \dots, n.$$

Since the X_i's are independent, if we let $Z = \sum_{i=1}^n X_i$, a well known property of mgf's makes it possible to write

$$M_Z(t) = M_{X_1}(t) \times M_{X_2}(t) \times \dots M_{X_n}(t).$$

As the X_i's are also identically distributed, we can write:

$$M_Z(t) = [M_{X_1}(t)]^n = \left(\frac{1}{1 - \lambda t}\right)^n$$

and this holds $\forall t < 1/\lambda$. This is easily recognized to be the mgf of a gamma(n, λ) random variable.

To prove (b), one can either use an mgf technique or a technique based on transformation of random variables. The technique based on moment generating functions is probably less cumbersome. To this purpose let X_1 and X_2 be two independent random variables both with an exponential(λ) distributions and let $Z = X_1 - X_2$. Thus, we can write

$$M_Z(t) = M_{X_1 + (-1) \cdot X_2}(t) = M_{X_1}(t) \cdot M_{X_2}(-t).$$

We know that $M_{X_1}(t) = (1 - \lambda t)^{-1}$ for $t < 1/\lambda$, so, all we need is computing $M_{X_2}(-t)$. Then,

$$M_{X_2}(-t) = \int_0^\infty e^{-tx} \frac{1}{\lambda} e^{-x/\lambda} dx = \int_0^\infty \frac{1}{\lambda} e^{-tx - x/\lambda} dx =$$

$$\frac{\lambda}{\lambda t + 1} \cdot \frac{1}{\lambda} \int_0^\infty \frac{\lambda t + 1}{\lambda} e^{-(t + 1/\lambda)x} dx = \frac{1}{1 + \lambda t} | - e^{-(t+1/\lambda)x}|_0^\infty = \frac{1}{1 + \lambda t}$$

which holds if and only if $t > -1/\lambda$. Therefore, we find that

$$M_Z(t) = M_{X_1}(t) M_{X_2}(-t) = \frac{1}{1 - \lambda t} \frac{1}{1 + \lambda t} = \frac{1}{1 - (\lambda t)^2} \;\; \forall t : |t| < 1/\lambda.$$

As this is the mgf for a double exponential$(0, \lambda)$, the second statement of the problem is proved.

To prove the last statement we can use basic facts about order statistics (see e.g., Casella and Berger, Ch. 5). The theorem states that if $X_1, X_2, \ldots, X_n$ are independent exponential(λ) random variables, and $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ denote the order statistics of the original sample then, $\min_i\{X_i\} = X_{(1)}$ and

$$f_{X_{(1)}} = n\left(\frac{1}{\lambda}e^{-x/\lambda}\right)\left(e^{-x/\lambda}\right)^{n-1} = \frac{n}{\lambda}e^{-nx/\lambda}$$

which is the pdf for an exponential(λ/n) random variable.

PROBLEM 2.2.2. X and Y are independent random variables with $X \sim \exp(\lambda)$ and $Y \sim \exp(\mu)$, respectively. It is impossible to obtain direct observations of X and Y. Instead, we observe the random variables Z and W, where

$$Z = \min\{X, Y\} \ W = \begin{cases} 1 & \text{if } Z = X; \\ 0 & \text{if } Z = Y. \end{cases}$$

(a) Find the joint distribution of Z and W.

(b) Prove that Z and W are independent.

SOLUTION. As X, Y are independent random variables, we have that

$$f_{X,Y}(x, y) = \frac{1}{\lambda\mu}\exp\{-x/\lambda - y/\mu\}.$$

The joint pdf of Z and W is given by:

$$f_{Z,W}(z, 1) = \frac{d}{dz}P(Z \le z, W = 1)$$

and

$$f_{Z,W}(z, 0) = \frac{d}{dz}P(Z \le z, W = 0).$$

Starting with the first of the two probabilities above, we see that

$$\begin{aligned} P(Z \le z, W = 1) &= P(X \le z, X \le Y) \\ &= P(0 \le X \le Y \le z) + P(0 \le X \le z \le Y). \end{aligned}$$

Hence, we have

$$\begin{aligned} P(Z \le z, W = 1) = &\int_z^\infty dy \int_0^z \frac{1}{\lambda\mu}\exp\{-x/\lambda - y/\mu\}dx \\ &+ \int_0^z dy \int_0^y \frac{1}{\lambda\mu}\exp\{-x/\lambda - y/\mu\}dx. \end{aligned}$$

A little calculus shows that

$$f_{Z,W}(z,1) = \frac{d}{dz}\left[\frac{\mu}{\mu+\lambda}(1-e^{-z\cdot\frac{\mu+\lambda}{\lambda\mu}}) = \frac{1}{\lambda}(1-e^{-z\cdot\frac{\mu+\lambda}{\lambda\mu}})\right]$$

for all $z \in [0,\infty)$.
In the same manner, we can prove that

$$f_{Z,W}(z,0) = \frac{1}{\lambda}(1-e^{-z\cdot\frac{\mu+\lambda}{\lambda\mu}})$$

for all $z \in [0,\infty)$. Using a compact notation, it is possible to write

$$f_{Z,W}(z,w) = \frac{w\mu+(1-w)\lambda}{\lambda\mu}\cdot(1-e^{-z\cdot\frac{\mu+\lambda}{\lambda\mu}})I[0,\infty)(z) \times I_{\{0,1\}}(w)$$

and since the joint pdf factors this suffices to assert that Z and W are independent random variables.
The other way to prove independence is that of proving that $P(Z \le z \mid W = 0,1) = P(Z \le z)$. In this case we have to compute the marginal pdf for W, which is given by

$$\begin{aligned} f_W(w) &= \int_0^\infty \frac{w\mu+(1-w)\lambda}{\lambda\mu}\cdot(1-e^{-z\cdot\frac{\mu+\lambda}{\lambda\mu}})I_{\{0,1\}}(w) \\ &= \frac{w\mu+(1-w)\lambda}{\lambda+\mu}. \end{aligned}$$

Thus,

$$\begin{aligned} f_{Z|W}(z) &= \frac{f_{Z,w}(z,w)}{f_W(w)} \\ &= \frac{\lambda+\mu}{\lambda\mu}\cdot(1-e^{-z\cdot\frac{\mu+\lambda}{\lambda\mu}})I_{[0,\infty)}(z). \end{aligned}$$

Since, this last pdf doesn't depend on w, it is clearly

$$P(Z \le z \mid W = 0,1) = P(Z \le z).$$

2.3 Transformations of Uniform Random Variables

PROBLEM 2.3.1. Find the pdf of $\Pi_{i=1}^n X_i$, where the X_i's are independent uniform(0, 1) random variables.
(Hint: Try to calculate the cdf, and remember the relationship between uniforms and exponentials.)

SOLUTION. The facts we need for this problem are summarized below without proofs:

Proposition 1. *If* $X \sim \text{uniform}(0,1)$, *then* $-\log(X)$ *is distributed as an exponential(1) random variable.*

Proposition 2. *If* $\{X_1, X_2, \dots, X_n\}$ *is a sequence of exponential(λ) random variables, then* $\sum_{i=1}^n X_i$ *is a gamma(n, λ) random variable.*

Now, if we let $Z = \Pi_{i=1}^n X_i$, we find that:

$$F_Z(z) = P(Z \le z) = P(\Pi_{i=1}^n X_i < z) = P(\sum_{i=1}^n -\log(X_i) > -\log(z))$$

for all $z \in (0,1]$, and using the analytical form of the pdf for $\sum_{i=1}^n -\log(X_i)$ (gamma(n, 1)) we get:

$$F_Z(z) = \int_{-\log(z)}^{\infty} \frac{1}{\Gamma(n)} x^{n-1} e^{-x} dx.$$

To find the pdf of Z it suffices to take the derivative of $F_Z(z)$ with respect to z. Using Leibnitz's rule we find

$$f_Z(z) = \frac{d}{dz} \int_{-\log(z)}^{\infty} \frac{1}{\Gamma(n)} x^{n-1} e^{-x} dx = \frac{1}{\Gamma(n)} \left(\frac{1}{\log(z)}\right)^{n-1}.$$

As this formula holds only for those $z \in (0,1]$ we are led to write:

$$f_Z(z) = \frac{1}{\Gamma(n)} \left(\frac{1}{\log(z)}\right)^{n-1} I_{(0,1]}(z).$$

PROBLEM 2.3.2. Let X and Y be two independent uniform(0, 1) random variables, and let Z = $X + Y$. Find the distribution of Z and its mgf.
Then let X, Y, and Z be three independent uniform(0, 1) random variables, let W = X + Y + Z, and find the distribution of W and its mgf.

SOLUTION. Even if the most efficient solution to this problem involves the use of convolutions, we will discuss it a solution using simple transformation rules. To this purpose, let $Z = X + Y$ and $W = X$. Then $Z \in [0,2]$ and $W \in [0,1]$ as it is easily seen, but this is not all of the problem. Using standard techniques on transformations of random variable we see that

$$f_{X,Y}(x,y) = 1 \cdot I_{[0,1]}(x) \cdot I_{[0,1]}(y);$$

$$X = W \text{ and } Y = Z - W, \Rightarrow \mid J \mid = 1;$$

$$f_{Z,W}(z,w) = 1 \cdot I_{[0,1]}(z) \cdot I_{[0,z]}(w) + 1 \cdot I_{(1,2]}(z) \cdot I_{[z-1,1]}(x).$$

The only thing that deserves being discussed is the form of the indicator functions. This happens because $W = X < X+Y = Z$ and, therefore, if $0 \leq Z \leq 1$, then $0 \leq W \leq z$, where z is the particular value taken by the random variable Z in $[0,1]$. In the same way when $Z \in (1,2]$ it is easily seen that, as $Y \in [0,1]$, we need $W = Y \in [z-1, 1]$ for Z to be in $(1,2]$.
Then, we have

$$f_Z(z) = \int_0^2 f_{Z,W}(z,w)dw = \begin{cases} z & \text{if } 0 \leq z \leq 1; \\ 2-z & \text{if } 1 < z \leq 2. \end{cases}$$

The plot of this pdf is given below.

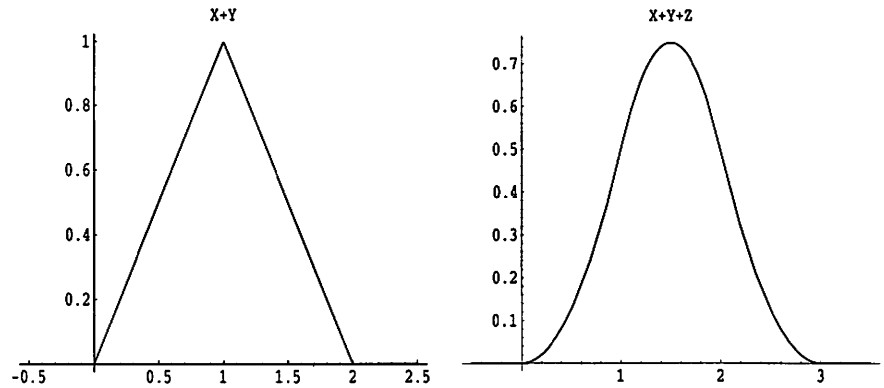

Figure 2.1: Plots of the pdfs for the sum of two and three uniform(0, 1) random variables, respectively.

The same problem arises when we consider differences, products and ratios of random variables. The pdf above is that of a triangular(0, 2) random variable. It is also simple to find the mgf for this pdf. In fact, using theorems on mgfs we have that for a uniform(0, 1) the mgf is given by $M_X(t) = (\mathrm{e}^t - 1)/t$, so $M_Z(t) = M_{X+Y}(t) = [M_X(t)]^2 = (\mathrm{e}^t - 1)^2/t^2$ which holds for all t in $\mathbf{R}$.

In the case of three i.i.d. uniform(0, 1) random variables, the problem is similar, though more cumbersome. We again address this problem using transformations. In particular the following transformations are used:

$$W = X + Y + Z; \quad V = Y; \quad U = X.$$

It is easily seen that the Jacobian for this transformation is 1. So, the joint pdf for W, V, and U is given by

$$f_{U,V,W}(u,v,w) =$$
$$I_{[0,1]}(w) \cdot I_{[0,w]}(v) \cdot I_{[0,w-v]}(u) + I_{(1,2]}(w) \cdot I_{[w-1,1]}(v) \cdot I_{[0,w-y]}(u)$$
$$+ I_{(1,2]}(w) \cdot I_{[0,w-1]}(v) \cdot I_{[w-1-y,1]}(u) I_{(2,3]}(w) \cdot I_{[w-2,1]}(v) \cdot I_{[w-2,1]}(u).$$

To see that these are the right indicator functions to use we have to use a plot for $X = U$, $Y = V$, and $X + Y = U + V$ for any given W reflecting on the fact that when $1 < W \leq 2$, for example, the relevant inequalities are

$$W - 1 \leq X + Y \leq W; \quad 0 \leq X \leq 1; \quad 0 \leq Y \leq 1.$$

In fact, if it were $X + Y < w - 1$ it would not be possible to attain $W = w \in (1, 2]$. When we look graphically we see that the subset of $[0,1] \times [0,1]$ satisfying these inequalities is the one represented in the second and third terms of the pdf for U, V, and W. The same procedure works for the other cases. Integrating out V and U gives:

$$f_W(w) = \begin{cases} w^2/2 & \text{if } 0 \leq w \leq 1; \\ -w^2 + 3w - 3/2 & \text{if } 1 < w \leq 2; \\ (3 - w)^2/2 & \text{if } 2 < w \leq 3. \end{cases}$$

The plot of this pdf was given in the figure above.
Using the same reasoning as for the case $X + Y$,we find that the mgf for $W = X + Y + Z$ is given by $(\mathrm{e}^t - 1)^3/t^3$ which holds for all $t \in \mathbf{R}$.

The discussion of the case of the sum of n independent U(0,1) random variables can be found in the book by W. Feller, An Introduction to Probability Theory and its Applications, Vol. II, 26-8. See also Problem 4.2.9 for a different approach.

2.4 Transformations of Normal Random Variables

PROBLEM 2.4.1. If X and Y are independent with density $f(x) = (2\pi)^{-1}e^{-x^2/2}$, show that $Z = X/Y$ has density $f(z) = \pi^{-1}(1 + z^2)^{-1}$.
Prove that $X/\mid Y \mid$ has also a Cauchy(0, 1) distribution.

SOLUTION. X, Y are i.i.d. $N(0,1)$, therefore the joint density function is given by

$$f_{X,Y}(x,y) = \frac{1}{2\pi}\exp\{-x^2/2 - y^2/2\}I_{\mathbf{R}}(x) \cdot I_{\mathbf{R}}(y).$$

Let $Z = X/Y$ and let $W = Y$. The transformation is well defined and 1-1 on $\mathbf{R}^2 - \mathcal{A}$, where $\mathcal{A} = \{\{x,y\} \in \mathbf{R}^2 : y = 0\}$. The fact that this transformation is not defined on $\mathcal{A}$ is not an important one. This is because the probability measure that induces the distribution function above (i.e.: $N(0,1) \times N(0,1)$) gives probability measure 0 to a set like $\mathcal{A}$. Then, if $\mathcal{S}$ is any event involving Z and W, we find that:

$$P[(Z,W) \in \mathcal{S}] = P[(X/Y,Y) \in \mathcal{S}] = P[(X/Y,Y) \in \mathcal{S} \cap \mathcal{A}^c] +$$

$$+ P[(X/Y,Y) \in \mathcal{S} \cap \mathcal{A}].$$

As $\mathcal{S} \cap \mathcal{A} \subset \mathcal{A}$ this implies that $P[\mathcal{S} \cap \mathcal{A}] = P[\mathcal{A}] = 0$, and therefore

$$P[(X/Y, Y) \in \mathcal{S}] = P[(X/Y, Y) \in \mathcal{S} \cap \mathcal{A}^c].$$

Then, if we like, we can define the transformation for those $(X, Y) \in \mathcal{A}$ in order to define it over the entire real plane and this will not change the probability for any event we consider and hence the distribution function is unchanged as well.
The general theory says that if $X_1, X_2, \dots, X_n$ are jointly continuous r.v.s with density function $f_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n)$, with $\mathcal{X} = \{(x_1, x_2, \dots, x_n) \in \mathbf{R}^n : f_{X_1,X_2,\dots,X_n}(x_1, x_2, \dots, x_n) \geq 0\}$, and $\mathcal{X}$ can be written as the partition of the sets $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_m$ such that $y_1 = g_1(x_1, x_2, \dots, x_n)$, $y_2 = (x_1, x_2, \dots, x_n)$, $\dots, y_m = g_m(x_1, x_2, \dots, x_n)$ are all 1-1 transformations from $\mathcal{X}_i$ onto $\mathcal{Y}$, $i = 1, 2, \dots, m$. Then, if $g_i(\cdot)$ is invertible on each $\mathcal{X}_j$, $i = 1, 2, \dots, n$, $i = 1, 2, \dots, m$, we find

$$f_{Y_1,Y_2,\dots,Y_n}(y_1, y_2, \dots, y_n) = \sum_{i=1}^{m} \mid J_i \mid f_{X_1,X_2,\dots,X_n}(g_{i_1}^{-1}(y_1, y_2, \dots y_n); \\ g_{i_2}^{-1}(y_1, y_2, \dots y_n); \dots g_{i_n}^{-1}(y_1, y_2, \dots y_n));$$

where $g_{i_j}^{-1}(y_1, y_2, \dots y_n) = x_i$ from $\mathcal{Y}$ to $\mathcal{X}_i$.
In our case, as our transformation is 1-1 on $\mathbf{R}^2 \cap \mathcal{A}$ we have (with some simple algebra):

$$f_{W,Z}(z, w) = \frac{1}{2\pi} e^{-w^2(1+z^2)/2} \mid w \mid \times I_{\mathcal{R}^2 \cap \mathcal{A}}(w, z).$$

As we are interested in the density for Z only, we have to integrate out W. This gives:

$$f_Z(z) = \int_{\mathcal{R}-0} f_{W,Z}(w, z) dw = \int_{\mathcal{R}} f_{W,Z}(w, z) dw$$

(having defined $f_{W,Z}(0, z) = 0$, for example)

$$\begin{aligned}
&= \int_{-\infty}^{0} \frac{\mid w \mid}{2\pi} e^{-w^2(1+z^2)/2} dw + \int_{0}^{\infty} \frac{\mid w \mid}{2\pi} e^{-w^2(1+z^2)/2} dw \\
&= \frac{1}{2\pi(1+z^2)} [e^{-w^2(1+z^2)/2} \mid_{-\infty}^{0}] + \frac{-1}{2\pi(1+z^2)} [e^{-w^2(1+z^2)/2} \mid_{0}^{\infty}] \\
&= \frac{2}{2\pi(1+z^2)} = \frac{1}{\pi(1+z^2)} I_{(-\infty,\infty)}(z)
\end{aligned}$$

which is the pdf for a Cauchy(0, 1) random variable.

To prove the second assertion we can use c.f.'s. It should be noted that when we do not know whether a r.v. has an mgf or not (and actually in our case it does not) we have to state our results in terms of c.f.'s. So, let $Z = X/Y$ and

$V = X/\mid Y \mid$. Then, using the law of total probability for expectations and the fact that X and Y are independent $N(0,1)$ random variables, we have

$$\begin{aligned}\psi_V(t) &= E_Y[E_{X|Y}[e^{it(X/|Y|)} \mid Y]] \\ &= E_Y[E_X[e^{it(X/|Y|)} \mid Y]] = E_Y[e^{-t^2/(2|Y|^2)}] = E_Y[e^{-t^2/(2Y^2)}] \\ &= E_Y[E_X[e^{it(X/Y)} \mid Y]] = E_Y[E_{X/Y}[e^{it(X/Y)} \mid Y]] = \psi_Z(t).\end{aligned}$$

Since $\psi_V(t) = \psi_Z(t)$, the Uniqueness Theorem for c.f.'s allows us to conclude that $X/\mid Y \mid$ is also a Cauchy(0, 1) random variable.

PROBLEM 2.4.2. Let $Y \sim N_p(\mu, \Sigma)$, $\mid \Sigma \mid \neq 0$, and let A be any $p \times p$ matrix of rank r. Then, $Y^TAY \sim \chi^2_{(r,\lambda)}$ iff $A\Sigma$ is an idempotent matrix.

SOLUTION. If we assume that the matrix $A\Sigma$ is idempotent and let $\Sigma = \Sigma^{1/2}\Sigma^{1/2}$ (this is possible since for a positive semi-definite matrix, M,we can always find a positive semi-definite matrix, N, such that $rank(N) = rank(M)$ and $M = N'N$), where $\mid \Sigma \mid \neq 0$. Then, if we define

$$\Omega = \Sigma^{1/2}A\Sigma^{1/2}$$

it is easily seen that:

i. Ω is idempotent. In fact,

$$\Omega\Omega = \Sigma^{1/2}A\Sigma^{1/2}\Sigma^{1/2}A\Sigma^{1/2} = \Sigma^{1/2}A\Sigma A\Sigma^{1/2}\Sigma^{1/2}\Sigma^{-1/2} =$$
$$\Sigma^{1/2}(A\Sigma)(A\Sigma)\Sigma^{-1/2} = \Sigma^{1/2}A\Sigma\Sigma^{-1/2} = \Sigma^{1/2}A\Sigma^{1/2} = \Omega,$$

using the assumption that $A\Sigma$ is idempotent and the fact that $\Sigma^{-1/2}$ exists as $\mid \Sigma \mid \neq 0$;

ii. rank$(\Omega) = r$. In fact, rank$(A) = r$ and, by assumption, Σ is not singular, hence $\Sigma^{1/2}$ is full rank.

The Spectral Decomposition Theorem, together with i.) and ii.), guarantees that it is always possible to find an orthogonal matrix Γ such that

$$\Gamma^T\Omega\Gamma = \begin{pmatrix} I_r & \emptyset \\ \emptyset & \emptyset \end{pmatrix}.$$

This implies that

$$\Gamma^T\Sigma^{1/2}A\Sigma^{1/2}\Gamma = \begin{pmatrix} I_r & \emptyset \\ \emptyset & \emptyset \end{pmatrix}$$

or

$$A = \Sigma^{-1/2}\Gamma \begin{pmatrix} I_r & \emptyset \\ \emptyset & \emptyset \end{pmatrix} \Gamma^T\Sigma^{-1/2}.$$

Letting $Z \equiv \Gamma^T \Sigma^{-1/2} Y$, we get that

$$Z \sim N_p(\Gamma^T\Sigma^{-1/2}\mu; \Gamma^T\Sigma^{-1/2}\Sigma\Sigma^{-1/2}\Gamma) = N_p(\Gamma^T\Sigma^{-1/2}\mu; I).$$

Now, using the expressions above, we can write

$$Y^TAY = (Y^T\Sigma^{-1/2}\Gamma)\begin{pmatrix} I_r & \emptyset \\ \emptyset & \emptyset \end{pmatrix} \cdot (\Gamma^T\Sigma^{-1/2}Y) = Z^T\begin{pmatrix} I_r & \emptyset \\ \emptyset & \emptyset \end{pmatrix} Z \equiv Z_1^T Z_1$$

where Z_1 is the vector of the first r-components of Z.
Since $Z_1 \sim N[(I_r \mid \emptyset)\Gamma^T\Sigma^{1/2}\mu; I]$ we have that

$$Z_1^T Z_1 \sim \chi^2_{(r,\lambda)}$$

where λ is given by[1]

$$\lambda = \frac{\mu^T[(I_r \mid \emptyset)\Gamma^T\Sigma^{1/2}]^T[(I_r \mid \emptyset)\Gamma^T\Sigma^{1/2}]\mu}{2} = \frac{\mu^T A\mu}{2}.$$

It is then easily seen that when $\mu = 0$ we have a central $\chi^2_{(r)}$.

Assume now that $Y^TAY \sim \chi^2_{(r,\lambda)}$. Since $Y^TAY = Z_1^TZ_1$ (according to what we proved before), this implies that $Z_1^TZ_1 \sim \chi^2_{(r,\lambda)}$. Therefore, if we take the m.g.f. of $Z_1^TZ_1$ it must equal that of a $\chi^2_{(r,\lambda)}$ r.v..(We could, of course, have used the c.f., but since the m.g.f. exists we can just use it.) For a non-central $\chi^2_{(r,\lambda)}$ random variable, the m.g.f. is given by:

$$m(t) = \left(\frac{1}{1-2t}\right)^{r/2} e^{-\lambda} e^{\frac{\lambda}{1-2t}}.$$

The moment generating function for Y^TAY is given by[2]

$$n(t) = \mid I - 2tA\Sigma \mid^{-1/2} \exp\{-\mu^T(I - (I - 2tA\Sigma)^{-1}\Sigma^{-1}\mu)/2\}.$$

So, we want $m(t) = n(t)$ and since we need this to be true for any choice of μ, we let $\mu = 0$. With this choice $m(t) = n(t)$ is the same as

$$(1-2t)^{r/2} = \mid I - 2tA\Sigma \mid^{1/2}$$

or

$$(1-2t)^r = \mid I - 2tA\Sigma \mid = \prod_{i=1}^{n}(1 - 2t\lambda_i)$$

[1]We are using the following theorem: If $Y \sim N_p(\mu, V)$, then $Z = Y'V^{-1}Y \sim \chi^2(p, \lambda)$ with $\lambda = \mu'V^{-1}\mu/2$.

[2]This relies on the following result which is easy to prove. If $Y \sim N_p(\mu, V)$, then the quadratic form $q = Y'AY$ has m.g.f.

$$m_q(t) = \mid I - 2tAV \mid^{-1/2} \exp\{-\mu'[I - (I - 2tAV)^{-1}]V^{-1}\mu/2\}$$

for $t < t_0$ the smalles root of the determinant in the expression above.

where the λ_i's are the eigenvalues of $A\Sigma$. In fact, if λ_i is an eigenvalue for $A\Sigma$, then it is a solution for the following

$$| A\Sigma - \lambda I | = 0$$

and hence $2t\lambda_i$ is a solution for

$$| 2tA\Sigma - \lambda I | = 0$$

and $| I - 2tA\Sigma | = \prod_i (1 - 2t\lambda_i)$. Equating $m(t)$ and $n(t)$ is the same as equating the coefficients of $m(t)$ to those of $n(t)$. It is easily seen that for $m(t)$ to be equal to $n(t)$ we need that r eigenvalues be equal to 1 and $p - r$ be equal to 0. Since this tells us that $A\Sigma$ has only eigenvalues 0's or 1's, we know that $A\Sigma$ is an idempotent matrix as we were suppose to show.

PROBLEM 2.4.3. On **R**, the double exponential distribution (DE(0,1)) is defined by the density $f_Z(z) = 1/2e^{-|z|}$. If X is DE(0, 1), prove that X can be represented as $Z_1 \cdot Z_2$ where $Z_2 \sim N(0,1)$, Z_1 and Z_2 independent and Z_1 is a positive random variable.

SOLUTION. The c.f. for a DE(0, 1) random variable, X, is given by

$$\psi_X(t) = \frac{1}{1+t^2}.$$

Assuming that there exist Z_1 and Z_2 as defined above such that $X = Z_1 \cdot Z_2$ we can write, using the law of total probability for expectations, that

$$\frac{1}{1+t^2} = E_{Z_1 \cdot Z_2}[e^{it(Z_1 Z_2)}]$$
$$E_{Z_1}[E_{Z_2|Z_1}[e^{it(Z_1 Z_2)} \mid Z_1]] = E_{Z_1}[E_{Z_2}[e^{it(Z_1 Z_2)} \mid Z_1]]$$

where in establishing the last equality the fact that Z_1 and Z_2 are independent was used. When Z_1 is given, the last expectation is easily seen to be

$$E_{Z_2}[e^{it(Z_1 Z_2)} \mid Z_1 = z_1] = e^{-(t^2 z_1^2)/2}.$$

Hence, since Z_1 is by assumption a positive random variable, we can rewrite the last expression found above as

$$\frac{1}{1+t^2} = \int_0^\infty f(z_1) e^{-(t^2 z_1^2)/2} dz_1$$

where $f(\cdot)$ is the pdf or the pmf for Z_1. Letting $z_1^2/2 = w$ it is possible to rewrite the last expression as

$$\frac{1}{1+t^2} = \int_0^\infty g(w) e^{-t^2 w} dw$$

with $g(w) = f(\sqrt{2w})/\sqrt{2w}$ the pdf or pmf for W. In addition, if we let $t^2 = s$, we can also write the basic equality as

$$\frac{1}{1+s} = \int_0^\infty g(w)e^{-sw}dw$$

and, since $\int_0^\infty g(w)e^{-sw}dw$ is the Laplace Transform for W, we have that W is a random variable whose Laplace Transform is given by $1/(1+s), s \geq 0$. The only random variable to satisfy this requirement is an exponential(1) random variable. This means that

$$W = \frac{Z_1^2}{2} \sim \text{exponential}(1)$$

and therefore, since $Z_1 = \sqrt{2W}$, a simple transformation gives that

$$f_{Z_1}(z_1) = z_1 e^{-z_1^2/2} I_{(0,\infty)}(z_1).$$

PROBLEM 2.4.4. Suppose that $X_1, X_2, \ldots, X_{N+M}$ are i.i.d. $N(\mu, \sigma^2)$ random variables, $N \geq 4$. Let

$$\bar{X} = (1/N)\sum_{i=1}^{N} X_i,$$

$$(N-1)s^2 = \sum_{i=1}^{N}(X_i - \bar{X})^2,$$

and

$$Y_i = (X_{N+i} - \bar{X})/\sqrt{(N-1)s^2}, \;\; i = 1, 2, \ldots, M.$$

(a) Derive the joint density of the random vector $Y' = (Y_1, Y_2, \ldots, Y_M)$.

(b) Determine the distribution of

$$\frac{N-1}{M} Y'\Omega^{-1}Y$$

where $\Omega = I_M + N^{-1}J_M$ and J_M is an $M \times M$ matrix all of whose elements are 1.

SOLUTION. If we let $Z' = (X_{N+1} - \bar{X}, \ldots, X_{N+M} - \bar{X})$, it is easily verified that

$$Z \sim N_M(0, \sigma^2\Omega)$$

where $\Omega = I_M + N^{-1}J_M$. In addition, it is a well know fact that

$$W \equiv (N-1)\frac{s^2}{\sigma^2} \sim \chi^2_{(N-1)}.$$

Since the random variable Y can be written as

$$Y = \frac{Z/\sigma}{\sqrt{W}},$$

we could use the simple rules for transformation of random variables to derive the joint density for the random vector Y. Nonetheless, here a different approach will be considered which requires a smaller amount of computations.
Precisely, the fact we use is the following:

$$f_{Y,W}(y,w) = f_{Y|W}(y) f_W(w).$$

It is easily seen that $Y \mid W = w \sim N_M(0, \Omega/w)$, therefore,

$$f_{Y|W}(y) = \frac{w^{M/2}}{(2\pi)^{M/2} \mid \Omega \mid^{1/2}} \exp\{-w/2(y'\Omega^{-1}y)\} I_{\mathbf{R}^M}(y).$$

Clearly, since $W \sim \chi^2_{(N-1)}$, we have also

$$f_W(w) = \frac{(1/2)^{(N-1)/2}}{w}^{(N-1)/2-1} \mathrm{e}^{-w/2} \Gamma(\frac{N-1}{2}) \times I_{[0,\infty)}(w).$$

Thus,

$$f_{Y,W}(y,w) = \frac{w^{(M+N-1)/2-1}}{(2\pi)^{M/2} \mid \Omega \mid^{1/2} \Gamma(\frac{N-1}{2}) 2^{(N-1)/2-1}}$$

$$\times \exp\{-w/2(1+y'\Omega^{-1}y)\} \times I_{\mathbf{R}^M}(y) I_{[0,\infty)}(w).$$

To derive $f_Y(y)$ it suffices to integrate out w from the joint pdf above. This gives

$$f_Y(y) = \int_0^\infty f_{Y,W}(y,w)dw =$$

$$\int_0^\infty \frac{w^{(M+N-1)/2-1}}{(2\pi)^{M/2} \mid \Omega \mid^{1/2} \Gamma(\frac{N-1}{2}) 2^{(N-1)/2-1}} \exp\{-w/2(1+y'\Omega^{-1}y)\} \times I_{\mathbf{R}^M}(y)dw =$$

$$\frac{1}{(2\pi)^{M/2} \mid \Omega \mid^{1/2} \Gamma(\frac{N-1}{2}) 2^{(N-1)/2-1}} \frac{\Gamma(\frac{M+N-1}{2})}{(1/2)^{(M+N-1)/2}(1+y'\Omega^{-1}y)^{(M+N-1)/2}} \times$$

$$\int_0^\infty \frac{w^{(M+N-1)/2}}{\Gamma(\frac{M+N-1}{2})} (\frac{1+y'\Omega^{-1}y}{2})^{(M+N-1)/2} \exp\{-(w/2)(1+y'\Omega^{-1}y)\} I_{\mathbf{R}^M}(y)dw =$$

$$\frac{1}{\pi^{M/2} \mid |\Omega \mid^{1/2} \Gamma(\frac{N-1}{2})} \cdot \frac{1}{(y'\Omega^{-1}y)^{(M+N-1)/2}} I_{\mathbf{R}^M}(y),$$

using the fact that the integrand function above is the kernel of a gamma($(M+N-1)/2$, $((1+y'\Omega^{-1}y)/2)^{-1}$) random variable.

To find the distribution of $T \equiv \frac{N-1}{M} Y'\Omega^{-1}Y$ it suffices to use the fact that $Y = \frac{Z/\sigma}{W^{1/2}}$ where Z and W were defined at the previous page. In this way, we can rewrite T as

$$T = \frac{N-1}{M} \frac{\frac{(\Omega^{-1/2}Z)'}{\sigma} \frac{(\Omega^{-1/2}Z)}{\sigma}}{W}.$$

Now,

$$\frac{(\Omega^{-1/2}Z)}{\sigma} \sim N_M(0, I)$$

and, therefore,

$$\frac{(\Omega^{-1/2}Z)'}{\sigma} \frac{(\Omega^{-1/2}Z)}{\sigma} \sim \chi^2_{(M)}.$$

Besides, we know from the definition of W that $W \sim \chi^2_{(N-1)}$. In addition, it is easily seen that these last two r.v.'s are independent. In fact, $W = g(s^2)$ and $(\Omega^{-1/2}Z) = f(\bar{X}, X_{N+1}, \ldots, X_{N+M})$ and s^2 is independent of $\bar{X}$ and X_{N+i} $i = 1, 2, \ldots, M$. Thus, it was established that

$$T = \frac{\chi^2_{(M)}/M}{\chi^2_{(N-1)}/(N-1)}$$

with $\chi^2_{(M)}$ independent of $\chi^2_{(N-1)}$. Thus,

$$T \sim F_{(M,N-1)}.$$

2.5 Transformations of Random Variables: General Problems

PROBLEM 2.5.1. Define $f_c : \mathbf{R}^n \mapsto [0, \infty)$ by

$$f_c(x) = \frac{C(n)}{(1 + x'x)^{(n+1)/2}}$$

where $C(n)$ is a constant, $n = 1, 2, \ldots$

(a) Show $\int f_c(x)dx < \infty$ and find the value of $C(n)$ which makes f_c a density. (Hint: Do it for $n = 1, 2$, and then try induction).

(b) Suppose $a \in \mathbf{R}^n$, $a \neq 0$. If X has density f_c, show that $Y = a'X$ has a scaled Cauchy distribution, that is $Y \stackrel{D}{=} cZ$ for some real c where Z has density $f(z) = \pi^{-1}(1 + z^2)^{-1}$ for $z \in \mathbf{R}$. (Hint: Show that ΓX and X have the same distribution whenever Γ is an orthogonal matrix and use this to show that it suffices to consider a of the form $(a_1, 0, \ldots, 0)'$.

SOLUTION[3]. Following the hint, we see that if $n = 1$ we have

$$\int_{-\infty}^{\infty} \frac{C(1)}{(1+x^2)}dx = C(1)\arctan(x) \mid_{-\infty}^{\infty} = C(1)\pi$$

which implies that if we let $C(1) = 1/\pi$ then f_c is a p.d.f. for some random variable (a Cauchy r.v. to be precise).
When $n = 2$, we have

$$\int_{\mathbf{R}^2} \frac{C(2)}{(1+x^2+y^2)^{3/2}}dxdy = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} \frac{C(2)}{(1+x^2+y^2)^{3/2}}dy$$
$$= \int_0^{\infty} dr \int_0^{2\pi} \frac{C(2)rdrd\theta}{(1+r^2)^{3/2}} = \frac{2\pi C(2)}{2}\int_0^{\infty} \frac{2r}{(1+r^2)^{3/2}}dr$$
$$= \pi C(2)\int_1^{\infty} \frac{dt}{t^{3/2}} = 2\pi C(2)[-t^{-1/2} \mid_1^{\infty}] = 2\pi C(2).$$

Hence, we need $C(2) = 1/2\pi$ in order for $f_c(\cdot)$ to be a p.d.f..
When $n = 3$,

$$\int_{\mathbf{R}^3} f_c(x,y,z)dxdydz = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{C(3)}{(1+x^2+y^2+z^2)^2}dxdydz.$$

Letting $y = r\cos(\theta)$ and $z = r\sin(\theta)$, we can rewrite the last integral in the form

$$\int_{-\infty}^{\infty} dx \int_0^{\infty}\int_0^{2\pi} \frac{C(3)rd\theta}{(1+x^2+r^2)^2} = \pi C(3)\int_{-\infty}^{\infty} dx \int_0^{\infty} \frac{2rdr}{(1+x^2+r^2)^2}$$
$$= \pi C(3)\int_{-\infty}^{\infty} dx \int_1^{\infty} \frac{dt}{(x^2+t^2)^2}$$
$$= \pi C(3)\int_{-\infty}^{\infty} -\frac{1}{x^2+t} \mid_1^{\infty} dx = \pi C(3)\int_{-\infty}^{\infty} \frac{1}{1+x^2}dx = \pi^2 C(3).$$

Therefore, in order for $f_c(\cdot)$ to be a p.d.f. we need this time $C(3) = 1/\pi^2)$. Repeating the same procedure also for the case $n = 4$ we find $C(4) = 3/(4\pi^2)$. This suggests the following choice for $C(i)$:

$$C(i) = \begin{cases} 1/\pi & \text{if } i = 1; \\ 1/(2\pi) & \text{if } i = 2 \\ (i-1)C(i-2)/(2\pi) & \text{if } i = 3, 4, \ldots, n. \end{cases}$$

We prove now that this formula is correct. We have to prove this in two steps: one when n is odd and one for n even. Consider first the case in which n is an odd number. In this case we can write n as $n = 2p + 1$ where $p = 0, 1, 2, \ldots$. Then,

$$\int_{\mathbf{R}^n} f_c(x_1, x_2, \ldots x_n)\Pi_{i=1}^n dx_i = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \cdots \int_{-\infty}^{\infty} \frac{C(n)dx_n}{(1+\sum_{i=1}^n x_i^2)^{p+1}}$$

[3]The present solution benefits from suggestions I received from Garrick Wallstrom.

and using once more polar transformations on x_{n-1} and x_n we get

$$\int_{\mathbf{R}^n} f_c(x_1, x_2, \dots x_n)\Pi_{i=1}^n dx_i =$$

$$= C(n)\pi \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \dots \int_{-\infty}^{\infty} dx_{n-2} \int_0^{\infty} \frac{2rdr}{(1+\sum_{i=1}^{n-2} x_i^2 + r^2)^{p+1}}$$

$$= C(n)\pi \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \dots \int_{-\infty}^{\infty} dx_{n-2} \int_1^{\infty} \frac{2rdr}{(\sum_{i=1}^{n-2} x_i^2 + r^2)^{p+1}} =$$

$$= \frac{C(n)\pi}{p+1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{dx_1 dx_2 \dots dx_{n-2}}{(1+\sum_{i=1}^{n-2} x_i^2 + r^2)^p} = \frac{C(n)\pi}{p+1} \cdot \frac{1}{C(n-2)}.$$

This implies that

$$\int_{\mathbf{R}^n} f_c(x_1, x_2, \dots x_n)\Pi_{i=1}^n dx_i = 1 \;\Rightarrow\; \frac{C(n)}{C(n-2)} \frac{\pi}{p} = 1$$

which means that $C(n) = (p/\pi)C(n-2)$ and, since $p = (n-1)/2$, we have

$$C(n) = \frac{(n-1)C(n-2)}{2\pi}$$

which agrees with what stated above.

When n is an even number, i.e. $n = 2p$, $p = 1, 2, \dots$, an analogous reasoning shows

$$\frac{C(n)\pi}{p-1/2} \cdot \frac{1}{C(n-2)} = 1$$

which implies

$$C(n) = \frac{C(n-2)}{\pi}(p-1/2) = \frac{(n-1)(n-2)}{2\pi}$$

using the fact that $p = n/2$. Hence, are done.

The usual way of finding the p.d.f. for $y = a'X$ doesn't work very well as we have:

$$Y \equiv \begin{pmatrix} a_{11} & \alpha' \\ \emptyset & I \end{pmatrix} X \equiv HX$$

where $\emptyset = (0, 0, \dots, 0)$, $\alpha' = (a_{12}, a_{13}, \dots a_{1n})$, $Y = (y_1, y_2, \dots, y_n)'$ and $X = (x_1, x_2, \dots, x_n)'$. H is clearly nonsingular as $\mid H \mid = a_{11}$ and so we get

$$f_{Y_1, Y_2, \dots, Y_n}(y_1, y_2, \dots, y_n) = \frac{C(n)}{(1 + Y'(H^{-1})'H^{-1}Y)^{(n+1)/2)}} \frac{1}{\mid a_{11} \mid} =$$

$$= \frac{C(n)}{\mid a_{11} \mid (1 + Y'KY)^{(n+1)/2}}$$

where $K = (H^{-1})'H^{-1}$ is a symmetric matrix by the way it is defined. We want the p.d.f. for Y_1 only and therefore we need to integrate out $Y_2, Y_3, \dots, Y_n$. It is easy to check using the inverse of a partitioned matrix (see e.g. Magnus and Neudacker, Matrix Differential Calculus, page 11) that

$$H^{-1} = \begin{pmatrix} 1/(a_{11}) & -(1/a_{11})\alpha' \\ \emptyset & I, \end{pmatrix} \Rightarrow$$

$$K = (H^{-1})'H^{-1} = \begin{pmatrix} 1/(a_{11}^2) & -(1/a_{11}^2)\alpha' \\ -(1/a_{11}^2)\alpha' & I + (1/a_{11}^2)\alpha\alpha' \end{pmatrix}.$$

Even with this, the problem of integrating out the variables we don't need remains difficult to perform. We can use a different idea and, in fact, we prove that if $X \sim f_c(x)$ on $\mathbf{R}^n$, then $Y = \Gamma X$ is such that $Y \overset{D}{=} X$, whenever Γ is an orthogonal matrix. The proof is as follows: if $Y = \Gamma X$ then $X = \Gamma' Y$ (using the fact that Γ is an orthogonal matrix) and using standard theorems for transforming random variables, we find that

$$\| \partial X/\partial Y \| = \| \Gamma \| = 1$$

since $\| \Gamma \| = \Gamma'\Gamma = I$, and this implies that $1 = | \Gamma'\Gamma | = | \Gamma' | \cdot | \Gamma | = | \Gamma |^2$ therefore $| \Gamma | = \pm 1$. Hence we can write

$$\begin{aligned} f_Y(y) = f_c(\Gamma' y) \| \Gamma \| &= \frac{C(n)}{(1 + y'\Gamma\Gamma' y)^{(n+1)/2}} \\ &= \frac{C(n)}{(1 + y'Iy)^{(n+1)/2}} = \frac{C(n)}{(1 + y'y)^{(n+1)/2}}. \end{aligned}$$

which proves our statement. This fact is useful because once we write

$$Y = \begin{pmatrix} (a_{11}\alpha') \\ A_{(n-1)\times n} \end{pmatrix} X \equiv MX$$

where A is a suitable matrix consisting of $n-1$ independent vectors in $\mathbf{R}^n$ and independent also from the vector in the first line of M. In this way M turns out to be an $n \times n$ nonsingular matrix. For example, we can choose $A = [\emptyset_{(n-1)\times 1} \mid I_{(n-1)\times(n-1)}]$ as we did before. Using the Gram-Schmidt Orthogonalization we can always write M in a way that its rows form an orthogonal basis of $\mathbf{R}^n$. We can start from the first row which is left unchanged but for a scale factor. This is all we need to find the distribution of Y_1. In this way we get a diagonal matrix, not necessarily an orthonormal one. But, we know that it is always possible to transform this matrix to be orthonormal just renormalizing each of its rows through the modified Gram-Schmidt algorithm. Now let $M^\star$ be this last matrix. This gives

$$Y^\star = M^\star X \sim \frac{C(n)}{(1 + y^{\star\prime} y^\star)^{(n+1)/2}}.$$

We are interested in the distribution of y_1^* so we have to integrate out[4] the contribution of the other variables. Doing this we get

$$f_{Y_1^*}(y_1) \sim \frac{C(1)}{(1+y_1^{*2})} = \frac{1}{\pi(1+y_1^{*2})}.$$

We have to pay attention now to the fact that we were originally looking for the p.d.f. of $y_1 = a'X$, but we computed that of $y_1^* = (1/ \parallel a_{11}) \parallel a'X$. But, since $y_1 = \parallel a_{11} \parallel y_1^* = (a'a)^{1/2} y_1^*$ it is easily seen that

$$f_{Y_1}(y_1) = \frac{1}{\pi \parallel a_{11} \parallel} \frac{1}{\left(1 + \frac{y_1^2}{\parallel a_{11} \parallel^2}\right)}$$

which is exactly the result we were looking for, as this is the p.d.f. for a scaled Cauchy r.v. and proves that we need to consider only vectors of the form $(a_{11}, 0, \ldots, 0)$.

PROBLEM 2.5.2. Let X_1 and X_2 be two $\mathbf{R}^d$-valued random variables having probability distribution functions f and g, respectively. Let $H_2(f,g)$ be the Hellinger distance between f and g defined by

$$H_2(f,g) \equiv \left(\int_{\mathbf{R}^d} \left(\sqrt{f(\mathbf{x})} - \sqrt{g(\mathbf{x})}\right)^2 \, d\mathbf{x}\right)^{1/2}.$$

Assume that T is a transformation from $\mathbf{R}^d$ onto $A \subset \mathbf{R}^d$ such that T is a bijection of class C^1 whose inverse is also of class C^1 and assume that the determinant of the Jacobian of T^{-1}, $J_{T^{-1}}(\mathbf{x})$ does not vanish for any $\mathbf{x}$ in A. Then, let $Y_1 = T \cdot X_1$ and $Y_2 = T \cdot X_2$, and let $\tilde{f}$ and $\tilde{g}$ be their probability distribution functions. Show that $H_2(f,g) = H_2(\tilde{f}, \tilde{g})$.

SOLUTION. Let $\rho(f,g)$ be defined as follows:

$$H_2(f,g) \equiv \int_{\mathbf{R}^d} \sqrt{f(\mathbf{x}) \cdot g(\mathbf{x})} \, d\mathbf{x},$$

a quantity called Matusita affinity measure[5] between two probability density functions. Using the definition of ρ, it is easily seen that

$$H_2(f,g) = \Big(2[1 - \rho(f,g)]\Big)^{1/2}$$

[4] We can eliminate two variables at the time for $n \geq 3$ as done above. Then, depending on whether n is even or odd we are left with the case $n = 2$ or $n = 1$. In the second case we are done, in the first case we have to integrate out one more variable. This is done using the following rule for integration:

$$\int \frac{dx}{(a+bx^2)^{m+1}} = \frac{1}{2ma} \frac{x}{(a+bx^2)^m} + \int \frac{dx}{(a+bx^2)^m}.$$

In our case this formula specializes with $m = 1/2$, $a = 1 + x^2$ and $b = 1$.

[5] This quantity was introduced by Kakutani in 1948, but the Japanese mathematician Kameo Matusita was the first one to use it for statistical applications.

and, hence, proving that $H_2(f,g) = H_2(\tilde{f},\tilde{g})$ is the same as showing that $\rho(f,g) = \rho(\tilde{f},\tilde{g})$. To this purpose, under the assumptions of the problem, it is possible to use results about the probability element and the change of variable theorem for integrals (see e.g. F. Jones, Lebesgue Integration on Euclidean Space, Jones and Bartlett Publishers, 1993; Chapter 15.J) to write:

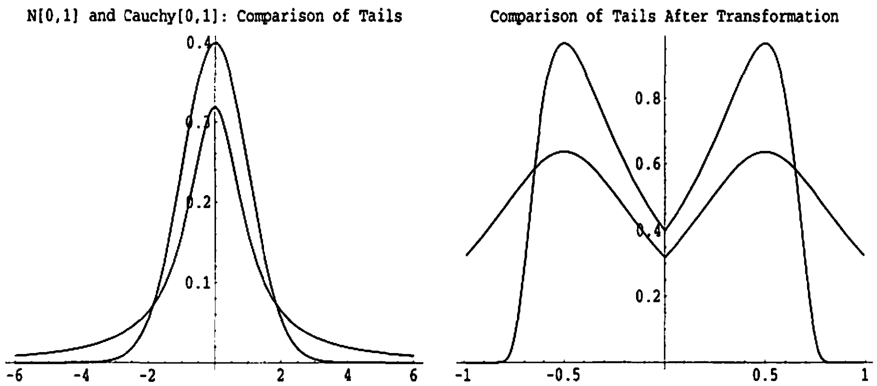

Figure 2.2: Distance between $N(0,1)$ and Cauchy$(0,1)$ random variables before and after the transformation $x \mapsto x/(1+|x|)$. The distance H_2 is the same and is approximately equal to .47. The probability mass hidden on the tails of the distributions shows clearly after the transformation.

$$
\begin{aligned}
\rho(\tilde{f},\tilde{g}) &= \int_A \sqrt{\tilde{f}(\mathbf{y})\cdot\tilde{g}(\mathbf{y})}\,dy \\
&= \int_{T^{-1}A} \sqrt{f(T^{-1}\mathbf{y})\cdot|J_{T^{-1}}(\mathbf{y})|\cdot g(T^{-1}\mathbf{y})\cdot|J_{T^{-1}}(\mathbf{y})|}\,dy \\
&= \int_{T^{-1}A} \sqrt{f(T^{-1}\mathbf{y})\cdot g(T^{-1}\mathbf{y})}\cdot|J_{T^{-1}}(\mathbf{y})|\,dy \\
&= \int_{\mathbf{R}^d} \sqrt{f(\mathbf{x})\cdot g(\mathbf{x})}\,d\mathbf{x} = \rho(f,g).
\end{aligned}
$$

PROBLEM 2.5.3. Let X and Y be two $\mathbf{R}^d$-valued random variables having probability distribution functions f and g, respectively. Let $H_2(f,g)$ and $L(f,g)$ be the distances between f and g defined by

$$
H_2(f,g) \equiv \left(\int_{\mathbf{R}^d} \left(\sqrt{f(\mathbf{x})}-\sqrt{g(\mathbf{x})}\right)^2\,d\mathbf{x}\right)^{1/2} \quad \text{and} \quad L(f,g) = \int_{\mathbf{R}^d} |f(\mathbf{x})-g(\mathbf{x})|\,d\mathbf{x},
$$

respectively. Prove the following inequalities:

(1) $L(f,g) \geq H_2^2(f,g)$;

(2) $L_1(f,g) \leq H_2(f,g)\sqrt{4-H_2^2(f,g)}$.

Combine (1) and (2) with the following LeCam's Inequality (LeCam, Convergence of Estimates Under Dimensionality Restrictions, Annals of Statistics, I pp. 38-53, (1973))

$$L_1(f,g) \leq 2 - [1 - \frac{1}{2}H_2^2(f,g)]^{1/2}$$

and discuss whether convergence in one metric is the same as convergence in the other metric.

SOLUTION. The first inequality can be proved as follows:

$$\begin{aligned} L(f,g) &= \int_{\mathbf{R}^d} |f(\mathbf{x}) - g(\mathbf{x})|\, d\mathbf{x} = \int_{\mathbf{R}^d} |\sqrt{f(\mathbf{x})} - \sqrt{g(\mathbf{x})}||\sqrt{f(\mathbf{x})} + \sqrt{g(\mathbf{x})}|\, d\mathbf{x} \\ &\geq \int_{\mathbf{R}^d} |\sqrt{f(\mathbf{x})} - \sqrt{g(\mathbf{x})}|^2\, d\mathbf{x} = H_2^2(f,g). \end{aligned}$$

The second inequality, on the other hand, can be verified using the Cauchy Schwarz inequality:

$$\begin{aligned} L^2(f,g) &= \left(\int_{\mathbf{R}^d} |f(\mathbf{x}) - g(\mathbf{x})|\, d\mathbf{x}\right)^2 \\ &= \left(\int_{\mathbf{R}^d} |\sqrt{f(\mathbf{x})} - \sqrt{g(\mathbf{x})}|\cdot|\sqrt{f(\mathbf{x})} + \sqrt{g(\mathbf{x})}|\, d\mathbf{x}\right)^2 \\ &\leq \int_{\mathbf{R}^d} (|\sqrt{f(\mathbf{x})} - \sqrt{g(\mathbf{x})}|)^2 \cdot \int_{\mathbf{R}^d} (|\sqrt{f(\mathbf{x})} + \sqrt{g(\mathbf{x})}|)^2 \\ &= H_2^2(f,g) \cdot [2 + 2\rho(f,g)] = H_2^2(f,g)[4 - H_2^2(f,g)] \end{aligned}$$

where $\rho(f,g)$ is the Matusita affinity measure between the two density functions f and g defined in Problem 2.5.2 and such that $\rho(f,g) = 2 - H_2^2(f,g)$. Taking the square root on both sides completes the proof.

Figure 2.3 shows clearly that convergence in H_2 implies convergence in L. However, the same figure also shows that the distance L can go to zero linearly or quadratically in H_2 and hence the speed of convergence can be different.

PROBLEM 2.5.4. Let $X_1, X_2, \ldots$ be independent, identically distributed random variables from the two parameter exponential distribution with density

$$f(x) = \frac{1}{\beta} e^{-(x-\alpha)/\beta} I_{[\alpha,\infty)}(x).$$

Show that

(a) $X_{(1)} \sim$ exponential$(\alpha, \beta/n)$;

(b) $X_{(i)} - X_{(i-1)} \sim$ exponential$(\beta/(n-i+1))$, $i = 2, 3, \ldots, n$;

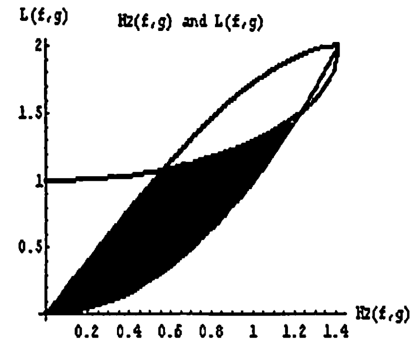

Figure 2.3: Relationship between Hellinger's H_2 and L distances. The shaded region represents the possible combination of $(H_2(f,g), L(f,g))$ for each couple of probability density functions, f and g, using inequalities (1), (2), and LeCam's.

(c) $X_{(1)}, X_{(2)} - X_{(1)}, \ldots, X_{(n)} - X_{(n-1)}$ are independent;

(d) $\sum_{i=1}^n X_i - nX_{(1)} \sim \frac{\beta}{2}\chi^2_{(2n-2)}$.

SOLUTION. The first three statements can be proved at the same time. This follows from being

$$g_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_1, x_2, \ldots x_n) = n!\Pi_{i=1}^n f_{X_i}(x_i)$$

where

$$f_{X_i}(x_i) = \frac{1}{\beta}e^{-(x_i-\alpha)/\beta} I_{[\alpha,\infty)}(x_i) \quad i = 1, 2, \ldots, n,$$

we find that

$$g_{X_{(1)},X_{(2)},\ldots,X_{(n)}}(x_1, x_2, \ldots x_n) = \frac{n}{\beta}e^{-(\sum_{i=1}^n x_i-\alpha)/\beta}\Pi_{i=1} I_{[\alpha,\infty)}(x_i).$$

If we introduce now the following transformations:

$$T_1 = X_{(1)}, T_2 = X_{(2)} - X_{(1)}, \ldots T_i = X_{(i)} - X_{(i-1)}$$

$i = 2, 3, \ldots, n$ it is easy to prove that the joint pdf for $T_1, T_2, \ldots T_n$ is given by

$$g_{T_1,T_2,\ldots,T_n}(t_1, t_2, \ldots t_n) =$$

$$\frac{n!}{\beta^n}e^{-(nt_1+(n-1)t_2+(n-2)t_3+\ldots 2t_{n-1}+t_n-n\alpha)/\beta} I_{[\alpha,\infty)}(t_1) \times \Pi_{i=2}^n I_{[0,\infty)}(t_i)$$

which can be rewritten as

$$g_{T_1,T_2,\ldots,T_n}(t_1, t_2, \ldots t_n) =$$

$$= \left[\frac{n}{\beta}e^{-n/\beta(t_1-\alpha)}\right] I_{[\alpha,\infty)}(t_1) \times \Pi_{i=2}^{n}\left[\frac{n-i+1}{\beta}e^{(n-i+1)t_i/\beta}\right] \times \Pi_{i=2}^{n} I_{[0,\infty)}(t_i).$$

Since the joint density factors in the product of the marginal densities of T_1, $T_2, \ldots T_n$, respectively, the independence of $X_{(1)}, X_{(2)} - X_{(1)}, \ldots, X_{(n)} - X_{(n-1)}$ follows. From the last expression above it is also easily seen that

$$X_{(1)} \sim \text{exponential}(\alpha, \beta/n)$$

and

$$X_{(i)} - X_{(i-1)} \sim \text{exponential}(\beta/(n-i+1)), \;\; i = 2, 3, \ldots, n.$$

To prove (d) it suffices to observe that

$$\sum_{i=1}^{n} X_i - nX_{(1)} = \sum_{i=1}^{n} X_{(i)} - nX_{(1)}$$

can be written in the following form

$$\sum_{i=2}^{n}(n-i+1)(X_{(i)} - X_{(i-1)}) = \frac{\beta}{2}\sum_{i=2}^{n}\frac{(n-i+1)}{2}(X_{(i)} - X_{(i-1)})$$

$$= \frac{\beta}{2}\sum_{i=2}^{n}\text{exponential}(2) = \frac{\beta}{2}\sum_{i=2}^{n}\text{gamma}(n-1, 2)$$

$$= \frac{\beta}{2}\sum_{i=2}^{n}\text{gamma}(2(n-1)/2, 2) = \frac{\beta}{2}\chi^2_{(2n-2)}$$

where the following facts were used:

Proposition 1. *If* $X \sim \text{exponential}(\beta)$ *then* $Y = X/\beta \sim \text{exponential}(1)$.

Proposition 2. *If* $X_1, X_2, \ldots X_n$ *are iid* exponential(α) *r.v.'s, then*

$$\sum_{i=1}^{n} X_i = \text{gamma}((n-1), \alpha)$$

.

Hence we found that

$$\sum_{i=1}^{n} X_i - nX_{(1)} \sim \frac{\beta}{2}\chi^2_{(2n-2)}$$

as required.

2.6 Non Elementary Distributions, Distributions of Stochastic Functionals and Stochastic Integrals

PROBLEM 2.6.1. [6] Let the joint density of X and Y be

$$f(x,y) = \frac{\pi^{1/2}\Gamma((m+n)/2)\Gamma((m+n-1)/2)}{\Gamma((n-1)/2)\Gamma(n/2)\Gamma((m-1)/2)\Gamma(m/2)} \cdot$$

$$\frac{xy^{(m-3)/2}(x-y)}{(1+x+y+xy)^{(m+n)/2}} I_{\{x \geq y \geq 0\}}(x,y).$$

Show that the random variable

$$Z = \frac{n-1}{m}[(1+X+Y+XY)^{1/2} - 1]$$

is distributed as a multiple of an F random variable with $2m$ and $2(n-1)$ degrees of freedom.

SOLUTION. This problem provides a good example of how direct transformation of random variables is not always the best way to proceed. In fact, if we were to use transformations like

$$Z = \frac{n-1}{m}[(1+X+Y+XY)^{1/2} - 1], \quad W = f(X,Y)$$

for some $f(\cdot)$ we will soon find that computations become very hard.
The following transformation:

$$\begin{cases} S = XY \\ T = 1 + X + Y + XY \quad ; \end{cases}$$

provides a good choice. In fact, it is $1-1$ in the region $\{(x,y) : x \geq y \geq 0\}$ and its Jacobian is given by $(x-y)^{-1}$. Therefore, we can write[7]

$$f_{S,T}(s,t) = \frac{\pi^{1/2}\Gamma((m+n)/2)\Gamma((m+n-1)/2)}{\Gamma((n-1)/2)\Gamma(n/2)\Gamma((m-1)/2)\Gamma(m/2)} \cdot$$

$$\frac{s^{(m-3)/2}}{t^{(m+n)/2}} I_{\{0 \leq (1+\sqrt{s})^2 \leq t\}}(s,t).$$

Defining now the new transformation

$$\begin{cases} Z = \frac{n-1}{m}(T^{1/2} - 1) \\ W = s \end{cases}$$

[6] This problem is due to Seymour Geisser.

[7] We know that as $(x-y)^2 \geq 0$ it must also be $x^2 + y^2 \geq 2\sqrt{xy}$, for any choice of x and y. Thus, $(\sqrt{x} - \sqrt{y})^2 \geq 0$ and, using the argument above, $x + y \geq 2\sqrt{xy}$. Now, by definition, $t = 1 + (x+y) + xy$ and $s = xy$. Using this fact and the inequality established above we are led to write $t \geq 1 + 2s^{1/2} + s = (1 + s^{1/2})^2$.

and using some simple algebra we can easily prove that

$$f_{W,Z}(w,z) = \frac{\pi^{1/2}\Gamma((m+n)/2)\Gamma((m+n-1)/2)}{\Gamma((n-1)/2)\Gamma(n/2)\Gamma((m-1)/2)\Gamma(m/2)}\times$$
$$\frac{2m}{n-1}\cdot\frac{1}{1+(\frac{mz}{n-1})^{m+n-1}}I_{\{0\le((n-1)\sqrt{w})/m\le z\}}(w,z).$$

Integrating out s we finally find

$$f_Z(z) = \frac{\pi^{1/2}\Gamma((m+n)/2)\Gamma((m+n-1)/2)}{\Gamma((n-1)/2)\Gamma(n/2)\Gamma((m-1)/2)\Gamma(m/2)}\times$$
$$\frac{4}{m-1}\frac{(\frac{mz}{n-1})^{m-1}}{(1+\frac{mz}{n-1})^{m+n-1}}I_{(0,\infty)}(z) =$$
$$\frac{\pi^{1/2}\Gamma((m+n)/2)\Gamma((m+n-1)/2)}{\Gamma((n-1)/2)\Gamma(n/2)\Gamma((m-1)/2)\Gamma(m/2)}\cdot\frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n-1)}\frac{4}{m-1}\times$$
$$\frac{\Gamma(m+n-1)}{\Gamma(m)\Gamma(n)}\left(\frac{m}{n-1}\right)^{2m/2}\frac{z^{2m/2-1}}{(1+\frac{mz}{n-1})^{(2m+2(n-1))/2}}I_{(0,\infty)}(z).$$

This last expression allows us to write

$$f_Z(z) = cF_{(2m,2(n-1))}$$

where

$$c = \frac{\pi^{1/2}\Gamma((m+n)/2)\Gamma((m+n-1)/2)}{\Gamma((n-1)/2)\Gamma(n/2)\Gamma((m-1)/2)\Gamma(m/2)}\cdot\frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n-1)}\frac{4}{m-1}$$

which is what we were supposed to prove.

PROBLEM 2.6.2. Let $W^0 = \{W^0(t) : 0 \le t \le 1\}$ be a Brownian bridge and let $W(t) = \{W(t) : 0 \le t \le 1\}$ be the associated Wiener process. Define the simple linear functional

$$L = \int_0^1 g(t)W^0(t)dt.$$

Show, stating any extra assumption it is thought to be necessary, that if $G(t) \equiv \int_0^t g(t)dt$, $0 \le t \le 1$, is square integrable , then

$$L \sim N(0,A), \text{ where } A = \int_0^1 G^2(t)dt - (\int_0^1 G(t)dt)^2.$$

Then, prove that

$$\int_0^1 g(t)W(t)dt \sim N(0,B) \text{ where } B = \int_0^1 G^2(t)dt + G^2(1) - 2G(1)\int_0^1 G(t)dt.$$

SOLUTION[8]**.** To prove normality for the linear functional L the reasoning is as follows. By definition, $W^0(t) = W(t) - tW(1)$, $0 \le t \le 1$, and it is easily proved that

$$(W^0(t_1), W^0(t_2), \dots W^0(t_m))' \sim N(\emptyset, \Gamma_m)$$

with

$$\Gamma_m = \begin{pmatrix} t_1 - t_1^2, & t_1 \wedge t_2 - t_1 t_2 & \dots & t_1 \wedge t_m - t_1 t_m \\ t_1 \wedge t_2 - t_1 t_2 & t_2 - t_2^2 & \dots & t_2 \wedge t_m - t_2 t_m \\ \dots & \dots & \dots & \dots \\ t_1 \wedge t_m - t_1 t_m & t_2 \wedge t_m - t_2 t_m & \dots & t_m - t_m^2. \end{pmatrix}$$

In addition, if we define

$$L_n = \sum_{i=0}^{n} g(i/n) W^0(i/n) \frac{1}{n},$$

it is easy to check that

$$L = \lim_{n \to \infty} L_n$$

and

$$L_n \sim N(\emptyset, H_n' \Gamma_n H_n)$$

with $H_n' = (g(0), g(1/n), \dots, g((n-1)/n), g(1))$. The following lemma completes the proof of normality.

Lemma 1. *Let X_n be a sequence of normal random variables converging in distribution to a random variable X. Then, X is either normal or constant.*

Proof. If we look at the sequence of c.f.'s associated with X_n, we have

$$\phi_{X_n}(t) = \exp\{it\mu_n - (1/2)t^2\sigma_n^2\},$$

where both μ_n and $\sigma_n^2 \not\to \infty$ as $n \to \infty$ or otherwise there is not convergence in distribution as tightness is lost. If $\mu_n \to \mu$ and $\sigma_n^2 \to \sigma^2 < \infty$, then

$$\phi_{X_n}(t) \to \phi_X(t) = \begin{cases} \exp\{it\mu - (1/2)\sigma^2 t^2\} & \text{if } \sigma^2 \neq 0; \\ \exp\{it\mu\} & \text{if } \sigma^2 = 0. \end{cases}$$

This completes the proof of the lemma and proves the first statement as well. □

The next task is proving that $E[L] = 0$. To do this we need a second lemma:

[8] Ron Pruitt and Charles Geyer contributed useful suggestions to the solution of this problem.

Lemma 2. *For $0 \leq t \leq 1$, let $Z_t(\omega)$ be a random variable defined on a probability space $(\Omega, \mathcal{F}, P)$. Assume that $Z_t \sim N(\mu t, \sigma^2 t)$ for some constant μ and σ^2 and also assume that $Z_t(\omega)$ is a continuous function of t for every $\omega \in \Omega$. Define:*

$$I(\omega) = \int_0^1 Z_t(\omega)dt \quad \omega \in \Omega.$$

Then,

1. *$Z_t(\omega)$ is a measurable function of (t, ω) on the product space $([0,1] \times \Omega, \mathcal{B} \otimes \mathcal{F}, \lambda \otimes P)$, where $\mathcal{B}$ is the Borel σ-field on $[0,1]$, λ is Lebesgue measure on $[0,1]$ and P is the probability measure associated with the assumption of a normal distribution for $Z_t(\omega)$.*

2. *$I(\omega)$ is a random variable and $E[I] = \mu/2$.*

Proof. See Problem 1.10.2 in Chapter 1. □

If we write $Z_t(\omega) = g(t)W_t^0(\omega)$ then, assuming that $g(t)$ is "nice", we can apply the second lemma to our integral. Using the fact that $W^0(t)$ is a Brownian Bridge, we have in this case that $Z_t \sim N(0, g^2(t)[t - t^2])$ and, thus, $E[L] = 0$. To compute the variance several integrations by parts are required. Assuming without loss of generality that $s < t$, we have

$$\begin{aligned}
E[L^2] &= E[\int_0^1 \int_0^1 g(s)g(t)W^0(s)W^0(t)dsdt] \\
&= 2E[\int_0^1 \int_s^1 g(s)g(t)W^0(s)W^0(t)dsdt] \\
&= 2\int_0^1 \int_s^1 g(s)g(t)E[W^0(s)W^0(t)]dsdt \\
&= 2\int_0^1 \int_s^1 g(s)g(t)s(1-t)dsdt
\end{aligned}$$

using again Fubini's Theorem and the fact that for a Brownian bridge

$$E[W^0(s)W^0(t)] = s \wedge t - st.$$

Letting $u = 1 - t$ and $dv = g(t)dt$, we have $du = -dt$ and $v = G(t)$ and, therefore, we get

$$\begin{aligned}
E[L^2] &= 2\int_0^1 sg(s)ds \int_s^1 g(t)(1-t)dt = \int_0^1 sg(s)ds\big[-G(t)(1-t)\,|_s^1 \\
&+ \int_s^1 G(t)dt\big] = \int_0^1 sg(s)\big[-G(s)(1-s) + \int_s^1 G(t)dt\big]ds \\
&= \int_0^1 -2s(1-s)g(s)G(s)ds + 2sg(s)\big[\int_s^1 G(t)dt\big]ds.
\end{aligned}$$

Now, let $u = -s(1-s)$, $dv = 2g(s)G(s)ds$ so that $du = (-1+2s)ds$ and $v = G^2(s)$. Then,

$$\int_0^1 -2s(1-s)g(s)G(s)ds = -2s(1-s)G^2(s)\mid_0^1 + \int_0^1 G^2(s)ds - \int_0^1 2sG^2(s)ds.$$

In addition, it is easily checked that

$$-2s(1-s)G^2(s)\mid_0^1 = 0$$

so that it is possible to write

$$E[L^2] = \int_0^1 G^2(s)ds - \int_0^1 2sG^2(s)ds + 2\int_0^1 sg(s)\Big[\int_s^1 G(t)dt\Big]ds.$$

The last partial integration involves $2\int_0^1 sg(s)[\int_s^1 G(t)dt]ds$. To this purpose, let $u = s$ and $dv = g(s)ds$. Then, $du = ds$ and $v = G(s)$. Thus, using Fubini's Theorem, we have

$$\begin{aligned}
\int_0^1 2sg(s)\Big[\int_s^1 G(t)dt\Big]ds &= \int_0^1\int_0^t g(s)sG(t)dsdt \\
&= 2\Big[\int_0^1 G(s)s\mid_0^t - \int_0^t G(s)ds\Big]G(t)dt = 2\int_0^1 \Big[G(t)t - \int_0^t G(s)ds\Big]G(t)dt \\
&= 2\int_0^1 G^2(t)tdt - 2\int_0^1\int_0^t G(s)G(t)dsdt = 2\int_0^1 G^2(t)tdt - \Big[\int_0^1 G(t)dt\Big]^2.
\end{aligned}$$

Thus,

$$\begin{aligned}
E[L^2] &= \int_0^1 G^2(s)ds - \int_0^1 2sG^2(s)ds + \int_0^1 2G^2(t)tdt - \Big[\int_0^1 G(t)dt\Big]^2 \\
&= \int_0^1 G^2(s)ds - \Big[\int_0^1 G(t)dt\Big]^2
\end{aligned}$$

as we were supposed to prove. In particular, because of the assumptions that G is square integrable, the variance of the limit distribution is finite.

To prove the second assertion, let $H = \int_0^1 g(t)W(t)$. Using the definition of Brownian bridge it is possible to write H as

$$H = L + W(1)\int_0^1 tg(t)dt$$

so that H is clearly normal being a linear combination of normal r.v.'s. To simplify the computations, we prove first the following fact.

Proposition. $\int_0^1 t(g(t)dt = G(1) - \int_0^1 G(t)dt$.

Proof. The proof requires a simple integration by parts. In fact, let $u = t$ and $dv = g(t)dt$, then we find that

$$\int_0^1 t(g(t)dt = tG(t) \mid_0^1 - \int_0^1 G(t)dt = G(1)1 - \int_0^1 G(t)dt.$$

□

Using this result it is easy to prove that $E[H] = 0$. To this purpose it suffices to write

$$E[H] = E[L] + E[W(1)][G(1) - \int_0^1 G(t)dt] = 0 + 0[G(1) - \int_0^1 G(t)dt] = 0.$$

Clearly, we have also

$$Var[H] = Var[L] + Var[W(1)G(1) - W(1)\int_0^1 G(t)dt] +$$

$$2Cov[L, W(1)G(1) - W(1)\int_0^1 G(t)dt].$$

$Var[L]$ was computed above. In addition, as $E[W(1)G(1) - W(1)\int_0^1 G(t)dt] = 0$, we have that $Var[W(1)G(1) - W(1)\int_0^1 G(t)dt] = E[W(1)G(1) - W(1)\int_0^1 G(t)dt]^2$. Hence,

$$E[W(1)G(1) - W(1)\int_0^1 G(t)dt]^2 = E[W^2(1)G^2(1) + W^2(1)[\int_0^1 G(t)dt]^2$$

$$-2W^2(1)G(1)\int_0^1 G(t)dt] = 1 \cdot G^2(1) + [\int_0^1 G(t)dt]^2 - 2G(1)\int_0^1 G(t)dt.$$

Finally,

$$Cov[L, W(1)G(1) - W(1)\int_0^1 G(t)dt]$$

$$= E[L \cdot W(1)[G(1) - \int_0^1 G(t)dt]] = [G(1) - \int_0^1 G(t)dt]E[LW(1)]$$

and, with simple manipulations, we find

$$E[LW(1)] = E[\int_0^1 g(t)W(t)W(1)dt - \int_0^1 g(t)tW(1)^2 dt]$$

$$= \int_0^1 g(t)E[W(t)W(1)]dt - \int_0^1 tg(t)E[W^2(1)]dt$$

$$= \int_0^1 g(t)E[W^2(t)]dt - \int_0^1 tg(t)E[W^2(1)]dt = \int_0^1 tg(t)dt - \int_0^1 tg(t) \cdot 1dt = 0$$

where the fact that for a Wiener process $E[W(t)W(s)] = E[W^2(s)] = s$, $s \leq t$, was used. Thus, the second statement of the problem follows easily.

PROBLEM 2.6.3. Let $W(t), -\infty < t < \infty$, be a Wiener Process with parameter σ^2. Let α and β be finite numbers and let g be a continuously differentiable function on $[\alpha, \beta]$. Find the distribution for the stochastic integral

$$\int_\alpha^\beta g(t)dW(t).$$

(Hint. Prove first that the usual integration by parts formula

$$\int_\alpha^\beta g(t)dW(t) = g(\beta)W(\beta) - g(\alpha)W(\alpha) - \int_\alpha^\beta g'(t)W(t)dt$$

holds.)

SOLUTION. In using the hint, one should notice that the integral $\int_\alpha^\beta g(t)dW(t)$ is not defined in the usual sense as the Wiener process $W(t)$ is nowhere differentiable. But, since $W(t)$ has continuous sample paths, one can write:

$$\int_\alpha^\beta g(t)dW(t) \equiv \lim_{\epsilon\to 0}\int_\alpha^\beta g(t)\left[\frac{W(t+\epsilon)-W(t)}{\epsilon}\right]dt.$$

Using integration by parts we have also that

$$\begin{aligned}
\lim_{\epsilon\to 0}\int_\alpha^\beta g(t)\left[\frac{W(t+\epsilon)-W(t)}{\epsilon}\right]dt &= \lim_{\epsilon\to 0}\int_\alpha^\beta g(t)\frac{d}{dt}\left(\frac{1}{\epsilon}\int_t^{t+\epsilon} W(s)ds\right)dt \\
&= f(t)\lim_{\epsilon\to 0}\int_t^{t+\epsilon} W(s)ds\ |_\alpha^\beta - \int_\alpha^\beta g'(t)\lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(\int_t^{t+\epsilon} W(s)ds\right)dt \\
&= g(t)W(t)\ |_\alpha^\beta - \int_\alpha^\beta g'(t)W(t)dt = \int_\alpha^\beta g(t)dW(t) \\
&= g(\beta)W(\beta) - g(\alpha)W(\alpha) - \int_\alpha^\beta g'(t)W(t)dt
\end{aligned}$$

as we were supposed to prove[9].
So, since the right hand side of the formula above is well defined and exists, we are done.

[9]If we let $G(t) = \int_\alpha^t W(s)ds$, $G(t+\epsilon) = \int_\alpha^{t+\epsilon} W(s)ds$ then, $G(t)$ and $G(t+\epsilon)$ are continuous as $W(s)$ is continuous and

$$W(t) = G'(t) = \lim_{\epsilon\to 0}\frac{G(t+\epsilon)-G(t)}{\epsilon} = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\int_t^{t+\epsilon} W(s)ds.$$

To find the distribution of $\int_g(t)dW(t)$ we can therefore use the equality just established and compute the distribution of

$$\int_\alpha^\beta g(t)dW(t) = g(\beta)W(\beta) - g(\alpha)W(\alpha) - \int_\alpha^\beta g'(t)W(t)dt.$$

Normality follows easily from basic properties of the Wiener process and from problem 15. Thus, we are left with having to compute the mean and the variance only. It is easy to check that the mean is zero. In fact:

$$E[\int_\alpha^\beta g(t)dW(t)] = g(\beta)E[W(\beta)] - g(\alpha)E[W(\alpha)] - \int_\alpha^\beta g'(t)E[W(t)]dt = 0$$

where the fact that since the integral exists, it is possible (by Fubini's Theorem) to interchange the operators E and $\int$ was used.
To compute the variance it is a little bit more complicated as it requires quite a little bit of algebra. Nevertheless all comes down to the following

Proposition. *Under the assumptions of the text of the problem and adding the assumption that h is another continuously differentiable function on $[\alpha, \beta]$, then*

$$E[\int_\alpha^\beta g(t)dW(t) \int_\alpha^\beta h(t)dW(t) = \sigma^2 \int_\alpha^\beta g(t)h(t)dt.$$

Proof. (See[10] Hoel, Port and Stone, Introduction to Stochastic Processes, 1972; pp. 142-4.) □

It is then easily seen from the fact above (let $h = g$) that

$$V[\int_\alpha^\beta g(t)dW(t)] = \sigma^2 \int_\alpha^\beta g^2(t)dt.$$

Hence, we have established that

$$\int_\alpha^\beta g(t)dW(t) \sim N[0, \sigma^2 \int_\alpha^\beta g^2(t)dt].$$

PROBLEM 2.6.4. Let $W(t), -\infty < t < \infty$, be a Wiener Process with parameter 1. Prove the following:

$$\int_0^1 W(t)dW(t) \sim \frac{1}{2}(\chi^2(1) - 1).$$

[10] If the reader wants to try proving this result by himself/herself, the first step is that of writing

$$\int_\alpha^\beta g(t)dW(t) = g(\beta)(W(\beta) - W(\alpha)) - \int_\alpha^\beta g'(t)[W(t) - W(\alpha)]dt$$

and do the same for the other stochastic integral.

SOLUTION. There are at least a couple of different ways to prove the statement of the problem.

1. Infinitesimal Calculus Approach. Since Wiener processes have continuous paths we can compute the integral using its definition. So, considering the more general stochastic integral, one can write

$$\int_0^t W(s)dW(s) = \lim_{n\to\infty}\sum_{k=1}^{n} W(\frac{k-1}{n}t)\,[W(\frac{k}{n}t)-W(\frac{k-1}{n}t)]$$

and, after simple algebraic manipulations,

$$= \frac{1}{2}W^2(t) - \lim_{n\to\infty}\frac{1}{2}\sum_{k=1}^{n}[W(\frac{k}{n}t)-W(\frac{k-1}{n}t)]^2$$

$$= \frac{1}{2}W^2(t) - \lim_{n\to\infty}\frac{t}{2n}\sum_{k=1}^{n}[\frac{\sqrt{n}}{\sqrt{t}}(W(\frac{k}{n}t)-W(\frac{k-1}{n}t))]^2.$$

Then, it is easily seen that

$$T_k \equiv \frac{\sqrt{n}}{\sqrt{t}}\left(W(\frac{k}{n}t)-W(\frac{k-1}{n}t)\right)$$

are i.i.d. $N(0,1)$ random variables. Thus, if one sets

$$S_n \equiv \frac{t}{2n}\sum_{k=1}^{n}\left[\frac{\sqrt{n}}{\sqrt{t}}\left(W(\frac{k}{n}t)-W(\frac{k-1}{n}t)\right)\right]^2$$

then,

$$S_n = \frac{t}{2n}\sum_{k=1}^{n}T_k^2.$$

Since $E[S_n] = t/2$ and $Var[S_n] = t/(2n)$, when $n \to \infty$, we have clearly

$$S_n \overset{q.m.}{\to} \frac{t}{2} \Rightarrow S_n \overset{D}{\to} \frac{t}{2}.$$

Thus, it has been established that

$$\int_0^t W(s)dW(s) \overset{q.m.}{\to} \frac{1}{2}(W^2(t) - t)$$

and letting $t = 1$ the conclusion follows[11].

2. Ito's Rule. A second way to prove this fact is provided by Ito's Rule[12]:

Ito's Rule. Let W be a Brownian Motion on $[0,1]$ and let $g : \mathbf{R} \mapsto \mathbf{R}$ be a twice continuously differentiable function. Then

$$g(W(t)) - g(W(0)) = \int_0^t g'(W(s))dW(s) + \frac{1}{2}\int_0^t g''(W(s))ds \quad a.s.$$

[11] $W(1)$ is a $N(0,1)$ random variable and convergence in quadratic mean implies convergence in distribution.

[12] See e.g. A. N. Shiryaev, Probability; 1996, pp 554-8.

If we let $g(W) = \frac{W^2}{2}$ then, $g'(W) = W$ and $g''(W) = 1$. Therefore,

$$\frac{1}{2}W^2(t) - 0 = \int_0^t W(s)dW(s) + \frac{1}{2}\int_0^t ds \quad a.s.$$

or

$$\int_0^t W(s)dW(s) = \frac{1}{2}[W^2(t) - t]$$

from which the statement of the problem can be easily derived.

PROBLEM 2.6.5. Let X be the stochastic process define by the following stochastic differential equation:

$$dX_t = \alpha X_t\, dt + \sigma X_t\, dW_t.$$

where W is Brownian motion. This defines a Geometric Brownian Motion with drift α and volatility σ. Prove that the stochastic process just defined has the following properties:

(1) If X starts at a positive value, it will remain positive. In particular, $X_t \to \infty$ a.s. as $t \to \infty$ if $\lambda > \sigma^2/2$, $X_t \to 0$ a.s. as $t \to \infty$ if $\lambda < \sigma^2/2$ and X_t keeps oscillating between 0 and ∞ as $t \to \infty$ if $\lambda = \sigma^2/2$.

(2) X has an absorbing barrier at 0.

(3) $X_t \mid X_s$ is lognormally distributed with mean $X_s \cdot e^{\alpha(t-s)}$ and variance $X_s^2 e^{2\alpha(t-s)} \cdot \left(e^{\sigma^2(t-s)} - 1\right)$. This shows that both the mean of the forecast tend to zero when $\alpha < 0$ and to infinity when $\alpha > 0$. Similarly, the variance of the forecast tends to infinity as t does when $\alpha \geq \sigma^2/2$ and to zero when $\alpha < \sigma^2/2$.

(4) The variance of the forecast $X_t \mid X_u$ tends to infinity as t does.

SOLUTION. Property (1) is not evident. But, if one assumes that X_0 is positive and solves the stochastic differential equation, the solution turns out to be:

$$X_t = X_0 \cdot \exp\{\alpha t + \sigma W_t\}$$

which is clearly always positive. Then, letting $Y_t = \log(X_t)$ it is easily seen (use Itô formula) that when $\sigma^2/2 \neq \lambda$

$$\lim_{t\to\infty} \frac{Y_t}{t} = \left(\alpha - \frac{\sigma^2}{2}\right) + \sigma \lim_{t\to\infty} \frac{W_t}{t} = \left(\alpha - \frac{\sigma^2}{2}\right)$$

as by the law of the iterated logarithm we know that $\lim_{t\to\infty} \sup \frac{W_t}{\sqrt{2t \log\log t}} = 1$ and $\lim_{t\to\infty} \inf \frac{W_t}{\sqrt{2t \log\log t}} = -1$. If $\sigma^2/2 = \lambda$, then it is

$$\lim_{t\to\infty} \sup \frac{Y_t}{t} = \sigma \quad \text{and} \quad \lim_{t\to\infty} \inf \frac{Y_t}{t} = -\sigma$$

from which the desired conclusions follow.
Property (2) does not require any comment. As for property (3), if one lets $Y_t = \ln(X_t)$, a simple application of Itô's formula gives that $dY_t = (\alpha - (1/2)\sigma^2)\,dt + \sigma\,dW_t$ from which it is easily seen that

$$Y_t \mid Y_s \sim N\Big(Y_s + (\alpha - \frac{\sigma^2}{2})(t-s), \sigma^2(t-s)\Big).$$

Finally, since $X_t = \exp\{Y_t\}$, we use the fact for which if a random variable $Z \sim N(\mu, \sigma^2)$ then $V = \exp\{Z\} \sim \text{Lognormal}(\mu, \sigma^2)$ and $E[V] = \exp\{\mu + \sigma^2/2\}$, $\text{Var}[V] = \exp\{2(\mu + \sigma^2)\} - \exp\{2\mu + \sigma^2\}$.

PROBLEM 2.6.6. Let X be a stochastic process defined by the stochastic differential equation below (mean reverting Ornstein-Uhlenbeck process):

$$dX_t = \alpha(X_t - m)\,dt + \sigma\,dW_t, \quad \alpha < 0.$$

Prove that:

(1) X maybe positive or negative.

(2) The conditional distribution of X_u given X_0 is normal with mean given by $m + (X_0 - m)e^{\alpha u}$ and standard error $\frac{\sigma}{\sqrt{-2\alpha}} \cdot \sqrt{(1 - e^{2\alpha u})}$.

(3) The variance of $X_u \mid X_0$ is increasing in u but tends to a finite number as u goes to infinity; the average of $X_u \mid X_0$, on the other hand is decreasing in u and tends to m as u goes to infinity.

SOLUTION. The normality of the distribution of X_u follows easily from the normality of $\int_0^u dW_s$. To compute the mean and variance of this distribution the following technical trick can be used. Let $X_u = Q_u e^{-u} + m$. Then, Itô's formula gives $dX_u = -Q_u e^{-u}\,du + e^{-u}\,dQ_u = (m - X_u)\,dt + e^{-u}\,dQ_u$. Since the drift terms match, one can compare the volatility terms, i.e. it must be $\sigma\,dW_u = e^{-u}\,dQ_u$ which implies that $dQ_u = \sigma e^u\,dW_u$. Simple algebraic manipulations give then

$$X_u = m + (X_0 - m)e^{-u} + \sigma \int_0^u e^{s-u}\,dW_s.$$

From this expression, it is easy to compute the mean and variance of the process. In fact, since $E[\int_0^u e^{s-u}\,dW_s] = 0$ and $E[\int_0^u e^{s-u}\,dW_s]^2 = \int_0^u e^{2(s-u)}\,dt$ using properties of Itô integrals.

PROBLEM 2.6.7. [13] Let $(X_1, X_2, \ldots)$ be and independent sequence of identically distributed $\bar{\mathbb{R}}$-valued random variables. Let

$$U(\omega) = \liminf_{n\to\infty} X_n(\omega) \;\; and \;\; V(\omega) = \limsup_{n\to\infty} X_n(\omega).$$

[13] This problem is from Gray and Fristedt's, *A Modern Approach to Probability Theory.*, op. cit. Garrick Wallstrom provided useful insights.

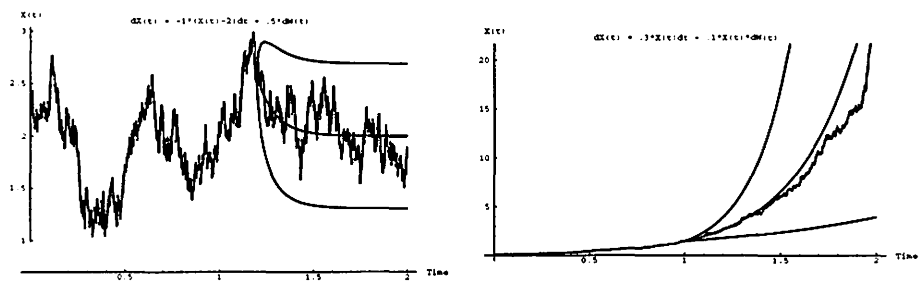

Figure 2.4: **Left.** *Mean reverting Oernstein-Uhlenbeck process with* $X_0 = 2.3$, $m = 2$, $\sigma = .5$ *and* $\alpha = -1.0$. *Superimposed to the plot are the mean function and the 95% pointwise confidence interval for* $X_u \mid X_t$ *when* $t = 1.2$ *and* $u \in (1.2, 2]$. **Right.** *Geometric Brownian motion with* $X_0 = .1$, $\alpha = .3$, $\sigma = .1$. *Superimposed to the plot are the mean function and the 95% pointwise confidence interval for* $X_u \mid X_t$ *when* $t = 1$ *and* $u \in (1, 2]$.

Find the distribution of the ordered pair (U, V) in terms of the common distribution function F of the X_n's. Is the pair (U, V) independent ?

SOLUTION. We need the following

Proposition. *Let* $(A_1, A_2, \ldots)$ *be an independent sequence of events. Then*

$$P[\cap_{n=1}^{\infty} A_n] = \Pi_{n=1}^{\infty} P(A_n).$$

Proof.

$$P[\cap_{n=1}^{\infty} A_n] = P[\lim_{m \to \infty} \cap_{n=1}^{m} A_n].$$

Let now $B_m \equiv \cap_{n=1}^{m} A_n$. Clearly, the B_m's form a decreasing sequence of events and hence the Continuity of Measure Theorem, together with the assumption of independence, implies that

$$P[\lim_{m \to \infty} \cap_{n=1}^{m} A_n] = \lim_{m \to \infty} P[\cap_{n=1}^{m} A_n] = \lim_{m \to \infty} \Pi_{n=1}^{m} P[A_n] = \Pi_{n=1}^{\infty} P[A_n].$$

□

By definition, $\forall a \in \bar{\mathbf{R}}$ and $\forall \epsilon > 0$, we have

$$\begin{aligned}F_V(a) &= P[\limsup_{n\to\infty} \leq a] \\ &= P[X_n \geq a + \epsilon, \text{for only finitely many } X_n's], \quad \forall \epsilon > 0 \\ &= P[\cap_{\epsilon>0} \cup_{n=1}^{\infty} \{\omega : X_n(\omega) < a + \epsilon, X_{n+1}(\omega) < a + \epsilon, \ldots\}] \\ &= P[\cap_{\epsilon>0} \cup_{n=1}^{\infty} \cap_{m=n}^{\infty} \{\omega : X_m(\omega) < a + \epsilon\}] \\ &= P[\cap_{r=1}^{\infty} \cup_{n=1}^{\infty} \cap_{m=n}^{\infty} \{\omega : X_m(\omega) < a + 1/r\}].\end{aligned}$$

Now, $\{\omega : X_m(\omega) < a + 1/r\}]$ forms a countable decreasing sequence of sets indexed by r and thus the Continuity of Measure Theorem implies

$$\begin{aligned}&\lim_{r\to\infty} P[\cup_{n=1}^{\infty} \cap_{m=n}^{\infty} \{\omega : X_m(\omega) < a + 1/r\}] \\ &\quad = \lim_{r\to\infty} \lim_{n\to\infty} P[\cap_{m=n}^{\infty} \{\omega : X_m(\omega) < a + 1/r\}] \\ &\quad = \lim_{r\to\infty} \lim_{n\to\infty} \Pi_{m=n}^{\infty} P[\{\omega : X_m(\omega) < a + 1/r\}]\end{aligned}$$

using the fact proved above and the assumption of independence for $X_1, X_2, \ldots$ This gives

$$F_V(a) = \lim_{r\to\infty} \lim_{n\to\infty} \Pi_{m=n}^{\infty} F(a + (1/r))^-.$$

Now, $\Pi_{m=n}^{\infty} F(a + (1/r))^-$ is either 0 or 1 for every choice of r, a and n. More precisely, we have

$$\lim_{n\to\infty} \Pi_{m=n}^{\infty} F(a + (1/r))^- = \begin{cases} 1 & \text{if } F(a + (1/r))^- = 1; \\ 0 & \text{if } F(a + (1/r))^- < 1. \end{cases}$$

Next, we should note that $\lim_{n\to\infty} \Pi_{m=n}^{\infty} F(a + (1/r))^-$ is a decreasing sequence indexed by r and therefore

$$\lim_{r\to\infty} \lim_{n\to\infty} \Pi_{m=n}^{\infty} F(a + (1/r))^- = \begin{cases} 1 & \text{if } F(a + (1/r))^- = 1; \\ 0 & \text{if } F(a + (1/r))^- < 1 \end{cases}$$

for any choice of r. Hence

$$F_V(a) = \begin{cases} 1 & \text{if } F(a) = 1; \\ 0 & \text{otherwise.} \end{cases}$$

Now, we have that $\forall a \in \bar{\mathbf{R}}$

$$\begin{aligned}F_U(a) &= P[\liminf_{n\to\infty} X_n(\omega) \leq a] \\ &= P[-\limsup_{n\to\infty}(-X_n(\omega)) \leq a] \\ &= P[\limsup_{n\to\infty}(-X_n(\omega)) \geq -a] \\ &= 1 - F_{\limsup(-X_n)}(-a)^-.\end{aligned}$$

Let $F_X \equiv F$, clearly it is $F_{-X}(a) = P[-X \leq a] = 1 - F_X(-a)^-$ and $F_{-X}(a)^- = 1 - F_X(a)$, so that, applying the result above, we have that

$$F_{\limsup(-X_n)}(-a) = \begin{cases} 1 & \text{if } F_{-X}(a)^- = 1; \\ 0 & \text{if } F_{-X}(-a)^- < 1 \end{cases} = \begin{cases} 1 & \text{if } 1 - F_X(a) = 1; \\ 0 & \text{if } 1 - F_X(a) < 1 \end{cases}$$
$$= \begin{cases} 1 & \text{if } F_X(a) = 0; \\ 0 & \text{if } F_X(a) > 0. \end{cases}$$

Thus, we found that $\forall a \in \bar{\mathbf{R}}$ it is

$$F_U(a) = \begin{cases} 1 & \text{if } F(a) = 0; \\ 0 & \text{if } F(a) > 0. \end{cases}$$

The two results just proved tell us that both U and V are constant random variables as both have a delta distribution. Thus (U, V) is independent as two constant random variables are always independent.

$$F_{U,V}(a,b) = P[\liminf_n X_n \leq a, \limsup_n X_n \leq b] \leq P[\liminf_n \leq a]$$

therefore, if $F_U(a) = 0$, then it is also $F_{U,V}(a,b) = 0$ for any $b \in \bar{\mathbf{R}}$. Similarly, $F_V(b) = 0$ then it is also $F_{U,V}(a,b) = 0$ for all $a \in \bar{\mathbf{R}}$. So the other only possibility is for $F_U(a) = F_V(b) = 1$. In this case

$$\begin{aligned} F_{U,V}(a,b) = P[U \leq a, V \leq b] &= P[U \leq a] + P[V \leq b] - P[U \leq a \text{ or } V \leq b] \\ &\geq P[U \leq a] + P[V \leq b] - 1 \end{aligned}$$

and, therefore,

$$F_{U,V}(a,b) = 1.$$

Thus, we found that for all $a, b \in \bar{\mathbf{R}}$ it is

$$F_{U,V}(a,b) = \begin{cases} 1 & \text{if } F(a) = 0, F(b) = 1; \\ 0 & \text{otherwise.} \end{cases}$$

PROBLEM 2.6.8. The Riemann zeta function is defined as

$$\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z}$$

for all complex z whose real part is greater than 1. Let $\mathbf{N} = \{1, 2, \ldots\}$ and let $\mathcal{F}$ be the collection of all subsets of $\mathbf{N}$. Prove that for each real $z > 1$, P_z, defined as

$$P_z(\{\omega\}) = \frac{1}{\zeta(z)\omega^z} \quad \omega \in \mathbf{N}$$

is a probability measure on the measurable space $(\mathbf{N}, \mathcal{F})$[14].
Compute the mean and the variance of a random variable, X having pmf P_z.
For each m calculate the probability of the event

$$\{\omega : \omega/m \in \mathbf{N}\}$$

and calculate the limit of this probability when $z \searrow 1$.
Let $\mathcal{P}$ be the set of prime numbers and consider for each ω in $\mathbf{N}$ the following expansion:

$$\omega = \Pi_{p\in\mathcal{P}} p^{X_p(\omega)}$$

where $X_p(\omega) \in \{0, 1, 2, \ldots\}$ denotes the power of the prime p in the prime factorization of ω. Prove that for the probability space $(\mathbf{N}, \mathcal{F}, P_z)$ the sequence $(X_p : p \in \mathcal{P})$ is independent. Also calculate and name the distribution for each X_p.

SOLUTION. Clearly $\mathcal{F}$ is a $\sigma-$field, therefore $(\mathbf{N}, \mathcal{F})$ is a measurable space. To prove that P_z as defined in the text is a probability measure we need to show that $P(\mathbf{N}) = 1$ and verify that P_z is countably additive. Both facts are easy to prove, in fact

$$P_z(\mathbf{N}) = \sum_{\omega=1}^{\infty} \frac{1}{\zeta(z)\omega^z} = \frac{\zeta(z)}{\zeta(z)} = 1$$

and, if $(A_1, A_2, \ldots)$ is a countable sequence of pairwise disjoint elements of $\mathcal{F}$ we have

$$P_z(\cup_{m=1}^{\infty} A_m) = \sum_{m=1}^{\infty} \sum_{\omega\in A_m} \frac{1}{\zeta(z)\omega^z} = \sum_{m=1}^{\infty} P(A_m)$$

which completes the proof of the first assertion.

The mean and the variance for a random variable X having P_z as pmf can be computed using the definitions. It is easily verified that

$$E[X] = \begin{cases} \zeta(z-1)/\zeta(z) & \text{if } z > 2; \\ \infty & \text{if } 1 < z \leq 2 \end{cases}$$

and

$$Var[X] = \begin{cases} \frac{\zeta(z-2)\zeta(z)-\zeta(z-1)^2}{\zeta(z)^2} & \text{if } z > 3; \\ \infty & if 2 < z \leq 3; \\ \text{undefined} & \text{if } 1 < z \leq 2. \end{cases}$$

To compute the probability of the event $\{\omega : \omega/m \in \mathbf{N}\}$ where $m \in \mathbf{N}$ it is necessary to calculate the following probability:

$$P[\{\omega : \omega = pm, p \in \mathbf{N}\} = \sum_{p=1}^{\infty} P_z(\{mp\})$$

$$\sum_{p=1}^{\infty} \frac{1}{\zeta(z)(mp)^z} = \frac{1}{m^z} \sum_{p=1}^{\infty} \frac{1}{\zeta(z)p^z} = \frac{1}{m^z}$$

[14]It is possible to think of P_z as being the distribution of the r.v. X defined as $X(\omega) = \omega$.

and, hence, when $z \searrow 1$ we get

$$P[\{\omega : \omega = pm, p \in \mathbf{N}\} \nearrow \frac{1}{m}.$$

If $\mathcal{P}$ is the set of the prime integer then, $\forall p \in \mathcal{P}$, using the fact that each $\omega \in \mathbf{N}$ has a unique expansion in terms of a product of prime numbers, we can write

$$P[\{\omega : X_p(\omega) \geq k\}] = \sum_{m=1}^{\infty} P[\{mp^k\}] = \sum_{m=1}^{\infty} \frac{1}{\zeta(z)(mp^k)^z}$$
$$\frac{1}{\zeta(z)p^{kz}} \sum_{m=1}^{\infty} \frac{1}{m^z} = \frac{\zeta(z)}{\zeta(z)p^{kz}} = \frac{1}{p^{kz}}.$$

This result does not depend on the particular choice of $p \in \mathcal{P}$ and, therefore, the distribution of X_p, $p \in \mathcal{P}$ is found to be geometric$(1 - 1/p^z)$. In fact, using the formula just proved for $P[X_p = k]$, we have that

$$f_X(x) = P[X_p = x] = P[X_p \geq x] - P[X_p \geq x+1] = \frac{1}{p^{kz}} - \frac{1}{p^{(k+1)z}}$$
$$= \left(\frac{1}{p^z}\right)^k \left(1 - \frac{1}{p^z}\right)$$

which proves our assertion.

We want to show now that the X_p's are independent. We do this using mathematical induction. Let's first consider the case of two r.v.'s, X_p and X_q :

$$P[X_p \geq k, X_q \geq h] = P[\{\omega : \omega = mp^k q^h; m = 1, 2, \ldots\}]$$
$$= \sum_{m=1}^{\infty} \frac{1}{\zeta(z)(mp^k q^h)^z} = \frac{1}{p^{kz} q^{hz}}.$$

Let $A = \{\omega : X_p(\omega) \geq k\}$ and $B = \{\omega : X_q(\omega) \geq h\}$. The formula proved above shows that we can write

$$P(AB) = P(A)P(B)$$

and few algebraic manipulations suffices[15] to prove that it is also

$$P[(AB)^c] = P(A^c)P(B^c)$$

for all choices of $A, B \in \mathcal{F}$. Thus, we have established that

$$F_{X_p,X_q}(k,h) = F_{X_p}(k)F_{X_q}(h)$$

[15]We can start proving that $P(A^cB) = P(A^c)P(B)$. To do this, we start from $P(B) = P[(A \cup A^c)B] = P(A^cB) + P(AB)$. But, we know that $P(AB) = P(A)P(B)$ and therefore, moving this term to the other side, we find $P(B)(1 - P(A)) = P(A^CB)$. Similarly, one proves that $P(AB^c) = P(A)P(B^c)$ and then, we can use these facts to prove that $P(A^cB^c) = P(A^c)P(B^c)$.

which proves that X_p and X_q are independent.
Assume now that $X_1, X_2, \ldots, X_n$ are independent, i.e.:

$$P[X_{p_1} \geq k_1, X_{p_2} \geq k_2, \ldots, X_{p_n} \geq k_n] = \Pi_{i=1}^{n} P[X_n \geq k_n]$$

and consider the case of the $n+1$ r.v.'s $X_1, X_2, \ldots, X_{n+1}$. It is easily seen that we have

$$P[X_{p_1} \geq k_1, X_{p_2} \geq k_2, \ldots, X_{p_{n+1}} \geq k_{n+1}] = \frac{1}{(\Pi_{i=1}^{n} p_i^{k_i}) p_{n+1}^{k_{n+1}}}$$

which proves that we can write again

$$P(AB) = P(A)P(B)$$

where B is the same set as above and A is defined this time as $A = \{\omega : X_{p_i}(\omega) \leq k_i, \forall i \in \mathrm{N}\}$. The same arguments used for the case of two random variables apply here to give

$$P[X_{p_1} \leq k_1, X_{p_2} \leq k_2, \ldots, X_{p_{n+1}} \leq k_{n+1}] = \Pi_{i=1}^{n+1} P[X_{p_i} \leq k_i].$$

This proves that

$$F_{X_{p_1}, X_{p_2}, \ldots, X_{p_{n+1}}}(k_1, k_2, \ldots, k_{n+1}) = \Pi_{i=1}^{n+1} F_{X_{p_i}}(k_i)$$

and, since n is arbitrary, we are done.

This problem is relevant as it shows that it is possible to define a sequence of independent random variables on a probability space that is not a product space.

2.7 Technicalities about Distribution Functions

PROBLEM 2.7.1. It is a well known fact that for $\mathbf{R}^1$-valued random variables that if the probability measure has no atoms (i.e. $P\{\omega : X(\omega) = x\} = 0 \ \forall x \in \mathbf{R}$) then the d.f. $F(x)$ is continuouos everywhere in $\mathbf{R}$.
Prove that there is not an analogous property in the multidimensional case.

SOLUTION. The following counterexample for an $\mathbf{R}^2$-valued random variable provides a proof for the statement of the problem [16]. For $0 < k < 1$, let

$$F(x,y) = \begin{cases} xy & \text{if } 0 \leq x \leq 1, \text{ and } 0 \leq y \leq k; \\ kx & \text{if } 0 \leq x < \infty \text{ and } k \leq y < \infty; \\ y & \text{if } 1 < x < \infty \text{ and } 0 \leq y \leq 1; \\ 1 & \text{if } x > 1 \text{ and } y > 1; \\ 0 & \text{otherwise.} \end{cases}$$

[16] This counterexample is a slight modification of one provided by J. M. Stoyanov in his book *Counterexamples in Probability; 1987*, p. 29.

In fact, it is easy tough tedious to verify that $F(x, y)$ just defined is a multidimensional distribution function, i.e.:

i. $F(x, y)$ is non decreasing in each of its arguments;

ii. $F(x, y)$ is right continuous in each of its arguments;

iii. $F(x, y) \to 0$ as $x \to -\infty$ or $y \to -\infty$;

iv. $F(x, y) \to 1$ as both $x, y \to \infty$;

v. If $a_i \leq b_i$, $i = 1, 2$, then

$$F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2) \geq 0.$$

In addition, each point $(x, y) \in T = \{(x, y) : 0 \leq x < \infty, 0 \leq y < \infty\}$ has zero probability. To check this it suffices to verify that $\forall (x, y) \in T$ it is

$$P[X = x, Y = y] = \lim_{h \searrow 0} P[X \in (x - h, x], Y \in (y - h, y]]$$

$$= \lim_{h \searrow 0} [F(x, y) - F(x - h, y)] - [F(x, y - h) - F(x - h, y - h)] = 0$$

using the fact that $F(x, y)$ is continuous in each of its arguments.

But, if one takes into consideration the set $S = \{(x, y) : k < y < \infty\}$, it is easily checked that

$$\lim_{h \searrow 0} F(1 + h, y + h) - F(1, y) = \begin{cases} y - k & \text{if } k < y < 1; \\ 1 - k & \text{if } y \geq 1. \end{cases}$$

Thus, $F(x, y)$ is discontinuous at each of the points in S.

PROBLEM 2.7.2. [17] A point x is said to belong to the support of the d.f. F iff for every $\epsilon > 0$, we have $F(x + \epsilon) - F(x - \epsilon) > 0$. The set of all such x is called the support of F.
Show that each point of jump belongs to the support, and that each isolated point of the support is a point of jump.
Give an example of a discrete d.f. whose support is the entire real line.

SOLUTION. If y is a point of jump for F then, $\forall \epsilon > 0$, it is possible to write:

$$F(y + \epsilon) - F(y - \epsilon) \geq F(y) - F(y-) > 0,$$

which, according to the definition, proves that y is in the support of F.

[17] This is problem 6, p. 10 in K. L. Chung's *A Course in Probability Theory*, 1974.

Let now z be an isolated point on the support of F. This means that the points in the set $\{y : | y - z |< \alpha\}$ do not belong to the support of F for all values of α sufficiently small. Thus, it is possible to find $\delta > 0$ such that

$$F(y + \delta) = F(y - \delta).$$

Now, since $F(\cdot)$ is a d.f., it is monotonically increasing and, therefore, it must be constant on $[y - \delta, y + \delta]$. This fact implies that F should be constant also on $(z - \alpha, z)$ and $(z, z + \alpha)$. Thus, if $\epsilon < \alpha$, it is possible to write that

$$0 < F(z + \epsilon) - F(z - \epsilon) = F(z+) - F(z-) = F(z) - F(z-)$$

proving that z is a point of jump.

For an example of a discrete random variable whose support is $\mathbf{R}$ one can consider the following:

$$F(x) = \sum_{n=1}^{\infty} 2^{-n}\delta_{\alpha_n}(x),$$

where $\{\alpha_n; n \geq 1\}$ is any given enumeration of the set of all rational numbers and

$$\delta_{\alpha_n}(x) = \begin{cases} 0 & \text{if } x < \alpha_n; \\ 1 & \text{if } x \geq \alpha_n. \end{cases}$$

In this case every point in $\mathbf{R}$ is in the support of $F(\cdot)$ as a consequence of the fact that $\mathbf{Q}$ is dense in $\mathbf{R}$.

PROBLEM 2.7.3. Let J_F be the set of discontinuities of a distribution function F and set $p(x) = F(x) - F(x-)$ for $x \in J_F$. Set $F_d(t) = \sum_{x \in J_F} p(x)\delta_x(t)$ for $t \in \mathbf{R}^1$.

(a) Show that $F_c(x) = F(x) - F_d(x)$ is nonnegative, continuous and nondecreasing on $\mathbf{R}^1$ with $F_c(-\infty) = 0$ and $F_c(\infty) \leq 1$.

(b) Using (a) show that every distribution function F can be written as $F = \alpha F_1 + (1 - \alpha)F_2$ where $\alpha \in [0, 1]$, F_1 is discrete and F_2 is continuous.

SOLUTION. We start by stating[18] and proving the following

Lemma. *Let F be d.f. with points of jump $\{a_i\}$. Then,*

$$\sum_{x-\epsilon<a_j<x} [F(a_j) - F(a_j-)] \to 0$$

and

$$\sum_{x-\epsilon<a_j le x} [F(a_j) - F(a_j-)] \to F(x) - F(x-)$$

as $\epsilon \searrow 0$ for all $x \in \mathbf{R}$.

[18] This lemma is Problem 2, p.9 in K. L. Chung, A Course in Probability Theory, 1974.

Proof. The series $\sum_{j=1}^{\infty}[F(a_j)-F(a_j-)]$ is clearly convergent and, therefore, for any choice of $\delta > 0$ it is possible to find $\bar{n}$ such that $\sum_{j>\bar{n}}[F(a_j)-F(a_j-)] < \delta$. If ϵ is chosen in such a way that $(x-\epsilon, x) \cap \{a_1, a_2, \dots, a_{\bar{n}}\} = \emptyset$, then one can write

$$\sum_{x-\epsilon<a_j<x} [F(a_j)-F(a_j-)] \le \sum_{j>\bar{n}}[F(a_j)-F(a_j-)] < \delta.$$

Since δ is arbitrary, this implies that $\sum_{x-\epsilon<a_j<x}[F(a_j)-F(a_j-)] \to 0$ as $\epsilon \searrow 0$. In the second case, we have that

$$\sum_{x-\epsilon<a_j\le x} [F(a_j)-F(a_j-)] =$$

$$\sum_{x-\epsilon<a_j<x} [F(a_j)-F(a_j-)] + F(x) - F(x-) \le \delta + F(x) - F(x-).$$

Again, since δ is arbitrary, the conclusion follows easily. □

Clearly, it is possible to write J_F as $J_F = \{x \in \mathbf{R} : F(x) - F(x-) > 0\}$. In addition, as a distribution function has at most countably many discontinuity points, it is possible to enumerate the elements of J_F.
We start proving that $F_c(x)$ is nonnegative. In fact,

$$F_d(x) = \sum_{x_i\in J_F : x_i \le x} p(x_i) = \sum_{x_i\in J_F : x_i \le x} F(x_i)-F(x_i-) \le F(x)-F(-\infty) = F(x).$$

This proves that $F_d(x) \le F(x)$, $\forall x \in \mathbf{R}$ and hence, using its definition, $F_c(x) \ge 0$.
The second proof concerns the fact that $F_c(x)$ is a non-decreasing function. To prove this, let $x_1 < x_2$ and note that according to its definition we have

$$F_c(x_2) - F_c(x_1) = [F(x_2) - F(x_1)] - [F_d(x_2) - F_d(x_1)].$$

The reasoning used above applies here as well. In fact, one can write:

$$F_d(x_2) - F_d(x_1) = \sum_{x_1<\alpha_i\le x_2} F(\alpha_i) - F(\alpha_i-) \le F(x_2) - F(x_1)$$

where it was chosen to take $x_1 < \alpha_i \le x_2$ because the point x_2 might belong to J_F. We have therefore established that $F_d(x_2) - F_d(x_1) \le F(x_2) - F(x_1)$, and hence $F_c(x_2) - F_c(x_1) \ge 0$ every time $x_1 < x_2$ as we were to show.
The third part of the proof concerns the continuity of $F_c(x)$. To prove it, as we are dealing with a situation in which there are points where $F(x)$ is only right continuous, it is more convenient to consider continuity from the right and continuity from the left, separately. As a matter of fact, the conclusion that $F_c(x)$ is continuous from the right at every $x \in \mathbf{R}$ follows from the fact that $F(x)$ being a distribution function is continuous from the right and $F_d(x)$ is also continuous from the right by construction. Then, as $F_c = F - F_d$, being a difference of two right continuous functions is also right continuous for every

$x \in \mathbf{R}$. Proving continuity from the left is a little bit more difficult than in the previous case. We want to show that $F_c(x) - F_c(x-\epsilon) \to 0$ as $\epsilon \searrow 0$. Now, we have:

$$F_c(x) - F_c(x-\epsilon) = [F(x) - F(x-\epsilon)] - [F_d(x) - F_d(x-\epsilon)] =$$

$$[F(x) - F(x-\epsilon)] - [\sum_{x-\epsilon<\alpha_i\leq x;\alpha_i\in J_F} F(\alpha_i) - F(\alpha_i-)].$$

Taking the limit for $\epsilon \searrow 0$, we find:

$$\lim_{\epsilon\searrow 0}[F_c(x) - F_c(x-\epsilon)] =$$

$$F(x) - \lim_{\epsilon\searrow 0} F(x-\epsilon) - \lim_{\epsilon\searrow 0}[\sum_{x-\epsilon<\alpha_i\leq x;\alpha_i\in J_F} F(\alpha_i) - F(\alpha_i-)] =$$

$$F(x) - F(x-) - [F(x) - F(x-)] = 0$$

using the second statement of the lemma above.
It is easy to prove that $F_c(-\infty) = 0$, and $F_c(\infty) = 1$. In fact, $F_c(-\infty) = F(-\infty) - F_d(-\infty) = 0 - 0 = 0$ and the first equality is therefore established. For the second one, we can start with the case in which F is a continuous distribution function. In this case $F_d(x) = 0\ \forall x \in \mathbf{R}$. Therefore $F_c(\infty) = F(\infty) - F_d(\infty) = 1 - 0 = 1$. When F is not continuous, then it has at least one jump and hence

$$F_d(\infty) = 1 \Rightarrow F_c(\infty) = 1 - F_d(\infty) < 1$$

and therefore we have that $F_c(\infty) \leq 1$. This completes the proof of part (a).

In the general case neither F_c nor F_d are distribution functions and precisely they fail in that both $F_c(\infty)$ and $F_d(\infty) \leq 1$. There is a simple way to transform them in order to have d.f.'s and precisely, let $\alpha = F_d(\infty)$, and write

$$F_1 = \frac{1}{\alpha}F_d,\quad F_2 = \frac{1}{1-\alpha}F_c$$

where of course it is $0 < \alpha < 1$. Now F_1 is a discrete distribution function while F_2 is a continuous distribution function. Multiplying F_d and F_c by a constant doesn't alter the properties proved above in part (a). Therefore, we can write the original distribution function F as

$$F(x) = \alpha F_1(x) + (1-\alpha)F_2(x) = F_d(x) + F_c(x)$$

everytime $0 < \alpha < 1$. When $\alpha = 0$, which implies that F is a discrete distribution function, we can just write $F(x) = F_d(x)$ or, in other words, we let $\alpha = 1$ in the general formula above. In a similar way, if F is a continuous d.f., then it must be $F(x) = F_c(x)$ which is the same as letting $\alpha = 0$ in the general formula. We have therefore shown that every d.f. F can be written as a convex combination of a discrete and a continuous d.f.'s.

PROBLEM 2.7.4. Suppose F and G are one dimensional distribution functions. Let $H_0(x,y) = \max\{F(x)+G(y)-1,0\}$ and $H_1(x,y) = \min\{F(x),G(y)\}$. Show that H_0 and H_1 are two dimensional distribution functions, each with F and G as marginal distributions. Further, show that if H is any other two dimensional distribution function with marginals F and G, then $H_0(x,y) \leq H(x,y) \leq H_1(x,y)$ for all x, y.
Using this example argue that there are infinitely many different distribution functions which have the same marginals.

SOLUTION. Let $F^{-1}(y) = \inf\{x : F(x) \geq y\}$ for any distribution function F. Now, let $U \sim \text{Uniform}(0,1)$ and define:

$$(X_1, X_2) = (F^{-1}(U), G^{-1}(u)).$$

We need the following facts whose proof can be found in most textbooks

Proposition 1. *If F is a distribution function and $U \sim$ Uniform$(0,1)$, then $X = F^{-1}(U) \sim F$.*

Proposition 2. *$F(F^{-1}(t)) \leq t$ where the inequality holds only if F is discontinuous at $F^{-1}(t)$.*

Then, using these two facts, we find that $F(x_1,x_2) = P[U \leq F(x_1), U \leq G(x_2)] = \min\{F(x_1), G(X_2)\}$. Hence H_1 is a bivariate distribution function.
Now, letting

$$(Y_1, Y_2) = (F^{-1}(U), G^{-1}(1-U)),$$

it is not difficult to check that we have

$$F(y_1,y_2) = P[U \leq F(y_1); 1-U \leq G(y_2)] = P[1-G(y_2) \leq U \leq F(y_1)].$$

It is also readily checked that this probability is $H_0(y_1,y_2)$ and hence H_0 is a bivariate distribution function.
For both of these distributions it is readily checked that the marginals are F and G. In fact, all it takes is evaluating the other argument at ∞, i.e.:

$$H_0(x,\infty) = \max\{F(x) + G(\infty) - 1)\} = F(x)$$
$$H_0(\infty,y) = \max\{F(\infty) + G(y) - 1)\} = G(y)$$

and similarly

$$H_1(x,\infty) = \min\{F(x), G(\infty)\} = F(x)$$
$$H_1(\infty,y) = \min\{F(\infty), G(y)\} = G(y)).$$

There is also the possibility to verify these facts using a direct approach (i.e.: verifying that both H_1 and H_0 satisfy the requirements to be distribution functions), but the calculations it requires are extremely heavy to perform, the

"rectangular property" in particular.
To check the last inequality, let $A = \{X \leq x\}$ and $B = \{Y \leq y\}$. Then $H(x,y) = P[A \cap B] \leq \min\{P(A), P(B)\}] = H_1(x,y)$. Besides, $H(x,y) = 1 - P[A^c \cup B^c] \geq 1 - P[A^c] - P[B^c] = P[A] + P[B] - 1$ and certainly $H(x,y) \geq 0$, so that $H(x,y) \geq H_0(x,y)$.
It is easily checked that if we write $H(x,y) \equiv \alpha H_0(x,y) + (1-\alpha)H_1(x,y)$, $0 < \alpha < 1$, then $H(x,y)$ has the same marginal distributions as $H_0(x,y)$ and $H_1(X,y)$ but, clearly, it is different from both $H_0(x,y)$ and $H_1(x,y)$.
This example can be further generalized to the case of probability distribution functions. See, for example, J. M. Stoyanov, Counterexamples in Probability, 1987; pp. 30-2.

PROBLEM 2.7.5. Give an example of a multivariate random variable, $Y' = (Y_1, Y_2, \ldots, Y_n)$ such that Y does not have a density function, but each of the Y_i's does.

SOLUTION. Let $Z \sim N(0, I_n)$ be a random variable in $\mathbf{R}^n$ and consider the new following random variable:

$$Y = \frac{Z}{\|Z\|}.$$

It is easily checked that with this definition it must be

$$Y'Y = \|Y\|^2 = 1.$$

Therefore, Y takes values on the $(n-1)$-sphere in $\mathbf{R}^n$ and, as such, this set has measure 0 with respect to the Lebesgue measure in $\mathbf{R}^n$.. Thus, Y cannot have a density.
Despite this, each of the Y_i's, where

$$Y_i = \frac{Z_i}{\| Z \|}, \quad i = 1, 2, \ldots, n$$

have a density. In fact, using the fact that the Z_i's are i.i.d $N(0,1)$ r.v.'s, it is possible to prove that

$$Y_i^2 \sim \text{Beta}(1/2, (n-1)/2) \quad i = 1, 2, \ldots, n.$$

This is a consequence of the fact that

$$Y_i^2 = \frac{Z_i^2}{\| Z \|^2} = \frac{Z_i^2}{Z_i^2 + \sum_{j=1, j \neq i} Z_j^2}, \quad i = 1, 2, \ldots, n$$

and, since $Z_i^2 \sim \chi^2(1)$, $\sum_{j=1, j\neq i} Z_j^2 \sim \chi^2(n-1)$ and Z_i^2 is independent of $\sum_{j=1, j\neq i} Z_j^2$, we are done[19]. It is now easy to prove that Y_i, $i = 1, 2, \ldots, n$ has

[19] The sum of $n-1$ independent $\chi^2(1)$ r.v.'s is a $\chi^2(n-1)$, or, a Gamma($(n-1)/2, 2$) r. v.. Then, use the fact that if $X \sim$ Gamma(α, 2) and $Y \sim$ Gamma(β, 2), we have that $X/(X+Y)$ is a Beta(α, β) random variable.

a density. A simple transformation of random variables suffices to show that its density is given by

$$f(y) = \frac{\Gamma(n/2)}{\Gamma(1/2)\Gamma((n-1)/2)}(1-y^2)^{(n-3)/2} I_{[-1,1]}(y).$$

Another example is the following: let $X \sim N(\mu; \Sigma)$ where $\Sigma = [1 + (1/n)]I_n - (1/n)J_n$ and $J_n = e_n e_n'$. In this case, the matrix Σ is singular and has rank $n-1$.[20] Now,

Proposition. *If we define the range of a linear transformation, $A : \mathbf{R}^n \mapsto \mathbf{R}^n$ as $\mathcal{R}(A) = \{u : u \in \mathbf{R}^n; u = Av \text{ for some } v \in \mathbf{R}^n\}$, then, $P(X \in \mathcal{R}(\Sigma) + \mu) = 1$.*

Proof. See Eaton, Multivariate Statistics, A Vector Space Approach., Proposition 2.7. □

Using this fact, we have that X belongs to a subset of $\mathbf{R}^{n-1}$ with probability one and therefore it doesn't have a density with respect to the Lebesgue measure. But, since rank$(\Sigma) = n-1$ every proper subset of X does have a density which is trivially seen to be the density for a multivariate normal in $\mathbf{R}^p$, $1 \leq p \leq n-1$.

PROBLEM 2.7.6. Let F be a distribution function. Prove that there is a sequence of discrete distribution functions $\{F_n\}$ such that:

$$\lim_{n\to\infty} \sup_x \mid F_n(x) - F(x) \mid = 0.$$

SOLUTION. If F is a discrete distribution function, that is,

$$F(x) = \sum_{i=1}^{\infty} p_i \delta_{x_i}(x),$$

with $\sum_{i=1}^{\infty} p_i = 1$ and

$$\delta_{x_i}(x) = \begin{cases} 1 & \text{if } x \geq x_i; \\ 0 & \text{if } x < x_i. \end{cases}$$

the problem is a trivial one since all we have to do is taking

$$F_n(x) = F(x), \; n = 1, 2, \ldots$$

Therefore, let's assume that F is a distribution function, but not a discrete one. This means that F can be continuous or with countably many discontinuity

[20] In fact, the matrix $(1 + 1/n)I_n - (1/n)J_n$ can be written as $\alpha P + \beta Q$ with $\alpha = 0$ and $\beta = (1 + 1/n)$. Then, the matrix $\alpha P + \beta Q$ has $n-1$ eigenvalues equal to β and one eigenvalue is α; therefore, since $n-1$ eigenvalues are not zero, the rank of this matrix is $n-1$. This fact is proved in Problem 2.8.1 at the end of this chapter.

points. We know that it must be $0 \leq F(x) \leq 1 \; \forall x \in \mathbf{R}$. Hence, given that $F(x)$ is bounded, if we take

$$d = 1/n; \;\; y_0 = 0; \;\; y_i = y_0 + d \cdot i; \;\; i = 1, 2, \ldots, n;$$

$$B_{n_i} = \{x \in \mathbf{R} : y_i \leq F(x) < y_{i+1}\}, \;\; i = 0, 1, 2, \ldots, n-1;$$

$$B_{nn} = \{x \in \mathbf{R} : F(x) \geq 1\};$$

we can define the new function:

$$F_n(x) \equiv \sum_{i=0}^{n-1} y_i \cdot I_{B_{ni}}(x) + 1 \cdot I_{B_{nn}}(x)$$

It is easily seen that $\{F_n(x)\}$ is a sequence of discrete distribution functions. In fact, if we define z_i to be the smallest number in B_{ni}, $i = 1, 2, \ldots, n$ we can write:

$$F_n(x) = \sum_{i=1}^{n} \tilde{y}_i \delta_{z_i}(x), \;\; \tilde{y}_i = 1/n, \;\; i = 1, 2, \ldots, n$$

and, of course, $\sum \tilde{y}_i = 1$. Besides, $F_n(-\infty) = 0$, $F_n(\infty) = 1$, $F_n(x)$ is non-decreasing $\forall x \in \mathbf{R}$, and besides, by the way the B_{ni}'s were selected, $F_n(x)$ is also right continuous. This proves that the function as defined above is actually a discrete distribution function. In addition, we have also that

$$\mid F_n(x) - F(x) \mid \leq 1/n \;\; \forall x \in \mathbf{R}.$$

In fact, if $x \in B_{ni}$, for some index i, $i = 1, 2, \ldots, n$, we have $F(x) \in [y_i, y_{i+1})$ and $F_n(x) = y_i$ which implies that $\mid F_n(x) - F(x) \leq \mid y_{i+1} - y_i \mid \leq 1/n$. Hence, $\sup_{x \in R} \mid F_n(x) - F(x) \leq 1/n$ and this implies that

$$\lim_{n \to \infty} \sup_x \mid F_n(x) - F(x) \mid \leq \lim_n 1/n = 0.$$

thus proving that $F_n(x) \to F(x)$ uniformly.
Note. In achieving uniform convergence, it is essential the property that $F(x)$ is bounded. In general, if $G(x)$ is any non-bounded function such that, for example, $G(x) \geq 0$, $\forall x \in \mathbf{R}$, we can define the sequence of functions:

$$G_n(x) = \sum_{i=1}^{n2^n} \frac{i-1}{2^n} I_{B_{ni}}(x) + n \cdot I_{C_n}(x),$$

where $B_{ni} = \{x \in \mathbf{R} : (i-1)/2^n \leq G(x) \leq i/2^n\}$ $n = 1, 2, \ldots, n2^n$, and $C_n = \{x \in \mathbf{R} : G(x) \geq n\}$. Then, we still have $G_n \to G$, but the convergence is no longer uniform.

PROBLEM 2.7.7. Prove that if $g(x,y)$, $(x,y) \in \mathbf{R}^2$, is a continuous two dimensional p.d.f. it doesn't follow necessarily that the corresponding marginals $g_1(x)$ and/or $g_2(y)$ are continuous functions as well. Show also that it is possible to find examples where $g(x,y)$ is a continuous two dimensional p.d.f. but at least one of the marginals have infinitely many discontinuities.

SOLUTION. To prove that continuous two dimensional p.d.f.'s do not always have continuous marginals it suffices to provide an example. To this purpose, let

$$g(x,y) = \frac{1}{4} \mid x \mid\mid y \mid \mathrm{e}^{-|x|-(1/2)|x|y^2} I_{(-\infty,\infty)}(x) \times I_{(-\infty,\infty)}(y).$$

It is easily checked that $g(x,y)$ is a p.d.f.; in fact, it is nonnegative everywhere in $\mathbf{R}^2$ and it integrates to 1. It is also easy to verify, starting from the definition, that $g(x,y)$ is continuous on $\mathbf{R}^2$.
The marginal density $g_1(x)$ is now given by:

$$g_1(x) = \begin{cases} (1/2)\mathrm{e}^{|x|} & \text{if } x \neq 0; \\ 0 & x = 0 \end{cases}$$

and clearly it has a discontinuity at $x = 0$. This completes the proof of the first part of the problem.
To find an example for the second one can consider the case where $g(x,y)$ is the p.d.f. defined above and set[21]

$$h(x,y) = \sum_{n=1}^{\infty} \frac{1}{2^n} g(x - r_n, y)$$

where $\{r_1, r_2, \ldots\}$ is an ordered infinite subset of the rational numbers. It is easy to check that $g(x - r_n, y)$ is a p.d.f. for all $n = 1, 2, \ldots$ and that such is $h(x,y)$. In addition, $h(x,y)$ is a continuous function but its marginal

$$h_1(x) = \sum_{n=1}^{\infty} g_1(x - r_n)$$

is discontinuous at $\{r_1, r_2, \ldots\}$, i.e., it has infinitely many discontinuities.

2.8 A Few Results Concerning an Important Matrix

PROBLEM 2.8.1. Suppose $n \geq 2$. Let $A_{a,b} = (a-b)I_n + bJ_n$ where I_n is an n dimensional identity matrix and $J_n = ee'$ where e is a length n vector of 1's. Let $P_e = n^{-1}ee' = n^{-1}J_n$ and $Q_e = I_n - P_e$.

(a) Prove that $\alpha P_e + \beta Q_e$ is non-singular iff $\alpha \neq 0$ and $\beta \neq 0$. Show that when $\alpha \neq 0$ and $\beta \neq 0$, $(\alpha P_e + \beta Q_e)^{-1} = \alpha^{-1}P_e + \beta^{-1}Q_e$.

[21] This idea is borrowed from J.M. Stoyanov, Counterexamples in Probability, 1987; pp. 32-3.

(b) Prove that $\alpha P_e + \beta Q_e$ is positive definite iff $\alpha > 0, \beta > 0$.

(c) Show that $A_{a,b}$ can be written as $\alpha P_e + \beta Q_e$ and give necessary and sufficient conditions on a and b so that $A_{a,b}$ is positive definite.

(d) Find $\det(A_{a,b})$.

SOLUTION. We start by stating and proving a few lemmas that will be useful in establishing our results.

Lemma 1. *If A is an $n \times n$ matrix with r, $(r \leq n)$ non zero eigenvalues, then rank$(A) \geq r$.*

Proof. See Magnus and Neudacker, Matrix Differential Calculus, Thm. 18. □

Lemma 2. *Let A be an $n \times n$ matrix with eigenvalues λ_i, $i = 1, 2, \dots, n$, then*

$$\text{tr}(A) = \sum_{i=1}^{n} \lambda_i \quad \text{and} \quad | A |= \Pi_{i=1}^{n} \lambda_i.$$

Proof. See Magnus and Neudacker, Thm. 17. Essentially, the two lemmas above are a direct consequence of Schur's Decomposition (see e.g. Bellman, Introduction to Matrix Analysis, 1970; Chapter 11, Thm. 4. □

Lemma 3. *Let x and y be two vectors in $\mathbf{R}^n$ different from the null vector. Then xy' has $n-1$ zero eigenvalues and one eigenvalue equals $x'y$.*

Proof. It should be easily seen that

$$xy' = \begin{pmatrix} x_1y_1 & x_1y_2 & \dots & x_1y_n \\ x_2y_1 & x_2y_2 & \dots & x_2y_n \\ \dots & \dots & \dots & \dots \\ x_ny_1 & x_ny_2 & \dots & x_ny_n \end{pmatrix}$$

It is also easily checked that one can write

$$(x_iy_1x_iy_2 \dots x_iy_n) = \frac{x_i}{x_1}(x_1y_1x_1y_2 \dots x_1y_n) \; i = 2, 3, \dots, n.$$

Therefore, the rank of xy' is 1 since we excluded the case $x = y = \emptyset$, the null vector. Then, using Lemma 1, we have that $x'y$ can have at most one non-zero eigenvalue and by Lemma 2 we also know that it is

$$x'y = \text{tr}(xy') = \sum_{i=1}^{n} \lambda_i$$

and, as only one eigenvalue can be different from zero, the proof is complete. □

Lemma 4. *Let $\lambda_i, i = 1, 2, \dots, n$ be the eigenvalues of the matrix A. Define $B = \delta I + \mu A$ and let $\gamma_i, i = 1, 2, \dots, n$ be its eigenvalues. Then*

$$\gamma_i = \delta + \mu\lambda_i, \quad i = 1, 2, \dots, n.$$

Proof. Using the definition of eigenvalue for a matrix B we have that for any x, an eigenvector of A, in V, (the space over which the matrix A is defined)

$$\gamma_i x = Bx = [\delta I + \mu A]x = \delta I x + \mu A x =$$
$$= \delta \cdot 1x + \mu\lambda_i x = (\delta + \mu\lambda_i)x$$

which proves our claim. Clearly, it is possible to write $\gamma_j = \delta + \mu\lambda_i$, $i \neq j$, but all this causes is a change of the order for $\gamma_i, i = 1, 2, \ldots, n$, but in any case this doesn't change their values. This is no longer true if B were defined as $B = \delta H + \mu A$ with $H \neq I$, unless all eigenvalues of H take the same value. Hence, this lemma is limited to the case $H = I$ or to the case in which the matrix H has an eigenvalues with multiplicity n. □

Lemma 5. $\mid \alpha I + \beta x y' \mid = \alpha^n + \alpha\beta x' y$.

Proof. Define $W = \alpha I + \beta x y'$. We know by lemma 4 that the eigenvalues of W are given by $\alpha + \beta\lambda_i(xy')$, $i = 1, 2, \ldots, n$, where $\lambda_i(xy')$ stays for the i-th eigenvalue of the matrix xy'. By lemma 3 we know that $n-1$ eigenvalues of xy' are zero and one equals $x'y$ and hence that $n-1$ of the eigenvalues of W are equal to α and one equals $\alpha + \beta x' y$. Lemma 2 is now enough to check that

$$\mid W \mid = \alpha^n + \alpha\beta x' y.$$

□

Lemma 6. *A symmetric $n \times n$ matrix A is positive definite iff all principal minors $\mid A_k \mid$, $k = 1, 2, \ldots, n$ are positive definite or, equivalently, iff the eigenvalues of A are all positive.*

Proof. For the first statement see Magnus and Neudacker, Thm. 29. For the second one see Cook, Larntz and Weisberg, Linear Models; 1993, Thm. 3.3.4. The equivalence is a direct consequence of the Spectral Theorem (or Schur's Decomposition). □

The six lemmas above provide us with the tools to provide quick answers to each of the statement we are asked to prove.

The matrix P_e can be written as xy' once we take

$$x' = y' = (1/n^{1/2}, 1/n^{1/2}, \ldots, 1/n^{1/2})$$

as it is easily checked. Then we have

$$\alpha P_e + \beta Q_e = \alpha P_e + \beta(I_n - P_e) = \beta I_n + (\alpha - \beta)P_e = \beta I_n + (\alpha - \beta)xx'.$$

By lemma 3 we know that the eigenvalues of xx' are $\lambda_i = 0, i = 1, 2, \ldots, n-1$ and $\lambda_n = x'x = 1$ (the order is not relevant). Using lemma 4 we find that the eigenvalues of $\beta I_n + (\alpha - \beta)xx'$ are given by

$$\mu_i = \begin{cases} \beta & \text{if } i = 1, 2, \ldots, n-1; \\ \alpha & \text{if } i = n \end{cases}$$

and thus, by lemma 2, we get

$$| \alpha P_e + \beta Q_e |=| \beta I_n + (\alpha - \beta)xx' |= \Pi_{i=1}^n \mu_i = \beta^{n-1}\alpha$$

which is different from zero iff $\alpha \neq 0$ and $\beta \neq 0$. To show that $(\alpha^{-1}P_e+\beta^{-1}Q_e) = (\alpha P_e + \beta Q_e)^{-1}$ when $\alpha, \beta \neq 0$ we need the following

Proposition. *P_e and Q_e are idempotent matrices besides, $P_eQ_e = Q_eP_e = \emptyset$, the $n \times n$ null matrix.*

Proof. We have already seen that it is possible to write

$$P_e = xx', \quad \text{with} \quad x' = (1/n^{1/2}, 1/n^{1/2}, \dots, 1/n^{1/2})$$

and thus

$$Pe \cdot P_e = xx'(xx') = x(x'x)x' = x \cdot 1 \cdot x' = xx' = P_e.$$

Similarly, using the result just established, we see that it is also

$$Q_e \cdot Q_e = (I - P_e)(I - P_e) = I - P_e - P_e + P_e^2 = I - P_e = Q_e$$

so,

$$P_e \cdot Q_e = P_e(I - P_e) = P_e - P_e^2 = P_e - P_e = \emptyset.$$

In the same way we can also prove that $Q_e \cdot P_e = \emptyset$.

□

Assuming now that $\alpha, \beta \neq 0$, we have that

$$\alpha^{-1}P_e + \beta^{-1}Q_e$$

is well defined and

$$(\alpha P_e + \beta Q_e) \cdot (\alpha^{-1}P_e + \beta^{-1}Q_e) =$$

$$= P_e + \alpha/\beta P_eQ_e + \beta/\alpha Q_eP_e + Q_e = P_e + \emptyset + \emptyset + Q_e = P_e + I - P_e = I$$

and in exactly the same way it is also possible to prove that

$$(\alpha^{-1}P_e + \beta^{-1}Q_e) \cdot (\alpha P_e + \beta Q_e) = I$$

thus proving, in accordance with the definition of inverse of a matrix, that

$$(\alpha^{-1}P_e + \beta^{-1}Q_e) = (\alpha P_e + \beta Q_e)^{-1}.$$

From lemma 4 we know that $(\alpha P_e+\beta Q_e)$ is positive definite iff its eigenvalues are all positive. We also saw that $(\alpha P_e + \beta Q_e)$ can be written as $(\beta I + (\alpha - \beta)xx')$, where x' was defined before. We have already proved that this last matrix has $n - 1$ eigenvalues equal to β and one eigenvalue equal to α. Therefore, if we want $(\alpha P_e + \beta Q_e)$ to be positive definite, we need $\alpha, \beta > 0$. Any other choice doesn't work. Thus, $(\alpha P_e + \beta Q_e)$ is positive definite iff $\alpha, \beta > 0$.

As $A_{a,b} = (a-b)I_n + bJ_n$ we can write it as $A_{a,b} = (a-b)I_n + nbP_e$ using the fact that $P_e = n^{-1}J_n$ and $A_{a,b} = (a-b)I_n + nbP_e + (a-b)P_e - (a-b)P_e = (nb + a - b)P_e + (a-b)[I_n - P_e] = (a + (n-1)b)P_e + (a-b)Q_e$ and this establishes that $A_{a,b} = \alpha P_e + \beta Q_e$ with $\alpha = a + (n-1)b$ and $\beta = a - b$.

We proved above that $\alpha P_e + \beta Q_e$ is positive definite iff both α and β are greater than 0. Thus a necessary and sufficient condition on a and b for $A_{a,b}$ to be positive definite is that

$$a + (n-1)b > 0, \quad a - b > 0$$

or

$$a > 0, \quad -a/(n-1) < b < a.$$

From lemma 2 we know that $\det(A_{a,b}) = \Pi_{i=1}^{n}\lambda_i(A_{a,b})$. We proved in part (c) that $A_{a,b}$ can be written as $(a + (n-1)b)P_e + (a-b)Q_e$ and we saw in part (a) that the eigenvalues for this last matrix are given by $\mu_i = a - b, i = 1, 2, \ldots, n-1$ and $\mu_n = a + (n-1)b$ (again, the order in which we write the eigenvalues is irrelevant). Hence

$$\det(A_{a,b}) = (a-b)^{n-1}(a + (n-1)b).$$

3 Chapter STOCHASTIC LIMIT THEOREMS

3.1 Monotone Convergence Theorem

PROBLEM 3.1.1. Let $X_1, X_2, \dots,$ be an increasing sequence of $\bar{\mathbf{R}}$-valued random variables on a common probability space $(\Omega, \mathcal{F}, P)$. For each $\omega \in \Omega$ set $X(\omega) = \lim_{n\to\infty} X_n(\omega)$.
Then, if $E[X_1] > -\infty$, it follows that $E[X_n] \to E[X]$ as $n \to \infty$.
Prove also that if $X_1, X_2, \dots,$ is a sequence of $\bar{\mathbf{R}}^+$-valued random variables, then

$$E[\sum_{n=1}^{\infty} X_n] = \sum_{n=1}^{\infty} E[X_n].$$

SOLUTION. As $X_1 \le X_2 \le X_3 \le \dots$, if we set $Y_1 = X_1$, $Y_n = X_n - X_1$; $n = 2, 3, \dots,$ and $Y = X - X_1$, then the Y_n's form an increasing sequence of positive random variables. A simple application of the MCT gives

$$E[Y_n] \nearrow E[Y] = E[X] - E[X_1].$$

Since $Y_n = X_n - X_1$ and $E[X_1] > -\infty$, $E[X_n - X_1]$ exists and $E[X_n - X_1] = E[X_n] - E[X_1]$ for all $n \ge 2$. The expression above can therefore be written as

$$E[X_n] - E[X_1] \nearrow E[X] - E[X_1].$$

If $E[X_1] < \infty$, it is then possible to add $E[X_1]$ to both terms. This gives

$$E[X_n] \nearrow E[X]$$

as $n \to \infty$. If, on the contrary, $E[X_1] = \infty$, then $E[X_n] = \infty$ $\forall n = 2, 3, \dots,$ and we have also $E[X] = \infty$. So, it is still true that $E[X_n] \nearrow E[X]$ as $n \to \infty$.

To prove the second statement, let $S_m = \sum_{n=1}^{m} X_n$. Then it is clearly $0 \le S_m \le S_{m+1}$ $\forall m = 1, 2, \dots$. Thus, one can use the MCT to state that

$$E[S_m] \nearrow E[S]$$

as $m \to \infty$, where $S \equiv \sum_{n=1}^{\infty} X_n$. Hence, as for finite sums of random variable it is always possible[1] to exchange the order of expectation and summation, we have established that

$$E[S] = \lim_{m\to\infty} E[S_m] = \lim_{m\to\infty} E[\sum_{n=1}^{m} X_n]$$

$$= \lim_{m\to\infty} \sum_{n=1}^{m} E[X_n] = \sum_{n=1}^{\infty} E[X_n]$$

which is the result we were supposed to prove.

PROBLEM 3.1.2. Let $X_1, X_2, X_3, \ldots,$ be a sequence of pairwise uncorrelated nonnegative random variables, and let $S = \sum_{n=1}^{\infty} X_n$. Assume that $E[S^2] < \infty$ and prove that

$$Var[S] = \sum_{n=1}^{\infty} Var[X_n].$$

Then, let $X_1, X_2, X_3, \ldots,$ be a sequence of random variables, all of which take values in the interval $[0,1]$. For each n, let F_n be the distribution function of X_n. Suppose that there is a right continuous F such that $\lim_n F_n(x) = F(x)$ exists for all x. Show that

$$\lim_{n\to\infty} E[X_n] = \int_0^{\infty} x dF(x).$$

SOLUTION. Let $S_m = \sum_{n=1}^{m} X_n$. Then, using the assumption according to which the X_i's are pairwise independent, one can write:

$$Var[S_m] = E[S_m^2] - (E[S_m])^2.$$

Since $S_1^2, S_2^2, S_3^2, \ldots,$ is a sequence of positive random variables such that $S_m^2 \nearrow S^2$, the MCT implies that

$$\lim_{m\to\infty} E[S_m^2] = E[S^2]$$

and a similar reasoning justifies the next conclusion:

$$\lim_{m\to\infty} E[S_m] = E[S].$$

Thus, using the definition of variance, it is possible to write that when $m \to \infty$

$$\sum_{n=1}^{m} Var[X_n] = \sum_{n=1}^{m} (E[X_n^2] - (E[X_n])^2) = E[(S_m)^2] - (E[S_m])^2$$

$$\nearrow E[S^2] - (E[S])^2 = Var[S]$$

[1] It is possible when the expectations are defined as it is the case when one is dealing with positive random variables.

which is well defined as, by assumption, $E[S^2] < \infty$ and, therefore, $E[S] < \infty$ as well. This completes the proof of the statement.

To prove the second statement, the reader should recall that for positive random variables it is $E[X] = \int_0^\infty (1 - F(x))dx$ (see Problem 1 in Chapter 1). Then, using the assumption about the convergence of $F_n(x)$ to $F(x)$ for all x and the fact that that $| 1 - F_n(x) | \leq 2$, for all n, we can use the BCT and interchange the operations of limit and integration. Thus,

$$\lim_{n\to\infty} E[X_n] = \lim_{n\to\infty} \int_0^1 (1 - F_n(x))dx$$

$$= \int_0^1 \lim_{n\to\infty} (1 - F_n(x))dx = \int_0^1 (1 - F(x))dx = \int_0^1 xdF(x) = \int_0^\infty xdF(x)$$

where integration by parts has been used.

PROBLEM 3.1.3. Calculate $E(X)$ and $E(X^2)$ for a random variable X having a geometric distribution by first calculating $E(X \wedge n)$ and $E((X \wedge n)^2)$ and then applying the Monotone Convergence Theorem as $n \to \infty$.

SOLUTION. If we define $Y_n = (X \wedge n)$ where X is a random variable with a geometric distribution, it is easily seen that $Y_1 \leq Y_2 \leq \dots$ form an increasing sequence of $\bar{\mathbf{R}}^+$-valued random variables on a common probability space $(\Omega, \mathcal{F}, P)$. It is also easily seen that $\forall \omega \in \Omega$ we have

$$X(\omega) = \lim_{n\to\infty} Y_n(\omega).$$

Then, using the MCT we have that

$$E[X] = \lim_n E[Y_n].$$

Therefore, to find the expected value of X we can first compute the expected value for Y_n for some fixed n and then we can take the limit for that last expression when $n \to \infty$. In this case we have:

$$E[Y_n] = \sum_{k=1}^{n-1} k(1-p)p^k + \sum_{i=n}^{\infty} n(1-p)p^i = p(1-p)\sum_{k=1}^{n-1} \frac{d}{dp}p^k + n(1-p)\sum_{i=n}^{\infty} p^i.$$

Since the first summation is a finite sum we can pull the derivative through the finite sum[2]. Simple algebra is now enough to write the last expression above as

[2]In general, when we deal with infinite sums we need some more. The general statement is as follows.
Suppose that the series $\sum_{x=0}^{\infty} h(p,x)$ converges for all $p \in (a,b)$ and $\frac{\partial h(p,x)}{\partial p}$ is continuous in p for each x and $\sum_{x=0}^{\infty} \frac{\partial h(p,x)}{\partial p}$ converges uniformly on every closed bounded subinterval of (a,b), then we can interchange summation and differentiation.

follows:

$$\begin{aligned}
E[Y_n] &= p(1-p)\frac{d}{dp}\sum_{k=0}^{n-1} p^k + n(1-p)\sum_{i=n}^{\infty} p^i \\
&= p(1-p)\frac{d}{dp}\Big(\frac{1-p^n}{1-p}\Big) + np^n(1-p) \\
&= \frac{p(p^n - np^{n-1} + 1 - p^n)}{1-p} + np^n(1-p).
\end{aligned}$$

When $n \nearrow \infty$ it is easy to check, by using the De L'Hopitâl's Rule, that:

$$E[X] = \lim_{n\to\infty} E[Y_n] = \frac{p}{1-p}.$$

To compute $E[X^2]$ the procedure is the same, i.e.:

$$\begin{aligned}
E[Y_n^2] &= (1-p)\frac{d^2}{dp^2}\sum_{x=1}^{n-1} p^x + n^2(1-p)\sum_{x=n}^{\infty} p^x \\
&= (1-p)\frac{(1-p)(-n(n-1)p^{n-2} + n^2p^{n-1} + 1 - (n+1)p^n)}{(1-p)^3} \\
&+ 2(1-p)\frac{(-np^{n-1} + np^n + p - p^{n+1})}{(1-p)^3} + n^2p^n(1-p)\frac{1}{1-p}.
\end{aligned}$$

It is easily seen that when $n \to \infty$ the last expression converges to

$$\frac{(1+p)p}{(1-p)^2},$$

thus proving that $E[X^2] = \lim_{n\to\infty} E[Y_n^2] = (p^2+p)/(1-p)^2$.

PROBLEM 3.1.4. Let $\Omega = (0,1)$ and let P be Lebesgue measure on the Borel subsets of $(0,1)$. Let X be an $\mathbf{R}$-valued random variable defined on Ω, and assume that $E[XI_{(a,b)}] = 0$ for all numbers $0 \le a, b \le 1$. Show that $X = 0$ a.s.
(Hint. First show that $E[XI_A] = 0$ for all events A.)

SOLUTION. Let $\mathcal{E} = \{(a,b) : 0 \le a < b \le 1\} \cup \emptyset$. In addition, let $\mathcal{D} = \{A : A \subset (0,1); E[XI_A] = 0\}$. The first conclusion is that $\mathcal{E} \subset \mathcal{D}$. It is also easily checked that $\mathcal{E}$ is closed under finite intersections. In fact,

$$(a_1,b_1)\cap(a_2,b_2) = \begin{cases} (\max\{a_1,a_2\}, \min\{b_1,b_2\}) & \text{if } \max\{a_1,a_2\} < \min\{b_1,b_2\}; \\ \emptyset & \text{otherwise.} \end{cases}$$

We prove now that $\mathcal{D}$ is a Dynkin class. To this purpose, it is easy to verify that

- $\Omega = (0,1) \in \mathcal{D}$;
- if $A_1, A_2 \in \mathcal{D}$ and $A_1 \subset A_2$ then $E[XI_{A_1-A_2}] = E[XI_{A_1} - XI_{A_2}] = E[XI_{A_1}] - E[XI_{A_2}] = 0 - 0 = 0$.

Assume now that $\{A_n : n = 1, 2, \ldots\}$ is a nondecreasing sequence of elements of $\mathcal{D}$. We want to show that $\cup_{n=1}^{\infty} A_n \in \mathcal{D}$. To this purpose, one should note that $E[XI_{A_n}] = 0$ for all $n = 1, 2, \ldots$, thus, it is possible to write that

$$0 = \lim_n E[XI_{A_n}] = \lim_n \Big(E[X^+ I_{A_n}] - E[X^- I_{A_n}] \Big) =$$

$$E[X^+ I_{\lim_n A_n}] - E[X^- I_{\lim_n A_n}] = E[X^+ I_{\cup_n A_n}] - E[X^- I_{\cup_n A_n}] = E[XI_{\cup_n A_n}].$$

The key step in the two lines above is that both $\{X^+ I_{A_n}, n = 1, 2, \ldots\}$ and $\{X^- I_{A_n}, n = 1, 2, \ldots\}$ are sequences of positive and nondecreasing functions such that $X^+ I_{A_n} \nearrow X^+ I_{\cup_n A_n}$ and $X^- I_{A_n} \nearrow X^- I_{\cup_n A_n}$. Interchanging the operation of limit and expectation is then just a consequence of applying the MCT. This completes the proof that $\mathcal{D}$ is a Dynkin class.
Now, as we have proved that $\mathcal{D}$ is a Dynkin class, $\mathcal{E} \subset \mathcal{D}$ and $\mathcal{E}$ is closed under finite intersections, Dynkin's $\pi - \lambda$ Theorem implies that $\sigma(\mathcal{E}) \subset \mathcal{D}$. The system of open intervals $\{(a, b) : 0 \leq a < b \leq 1\}$ generates $\mathcal{B}^{(0,1)}$, the Borel σ-field on $(0, 1)$, and therefore we have that $\sigma(\mathcal{E}) = \mathcal{B}^{(0,1)}$ and all events A are such that $E[XI_A] = 0$. If it were not true that $X = 0$ a.s., then there should exist a subset A of $(0, 1)$ such that $P(A) > 0$ and $E[XI_A] >$ or < 0. But, this would imply that $E[XI_A] \neq 0$. Hence a contradiction since we proved that for all events A it must be $E[XI_A] = 0$.

3.2 Dominated Convergence Theorem

PROBLEM 3.2.1. In the following, all r.v..'s are defined on the probability space $(\Omega, \mathcal{F}, P)$. Prove that for a random variable X with $E[|X|] < \infty$, and a sequence of events $A_n \in \mathcal{F}$ such that $A_n \downarrow \emptyset$, implies

$$\lim_n \int_{A_n} X dP = 0.$$

Then, prove that if $E[|Y|] < \infty$ and $\lim_n P(B_n) = 0$, it must be

$$\lim_n \int_{B_n} Y dP = 0.$$

SOLUTION. The first statement of the problem follows from an application of the Dominated Convergence Theorem (DCT). In fact, if one lets $Z_n = XI_{A_n}$ and $Z = XI_{\limsup_n A_n} = 0$, we have by assumption that

$$|Z_n| \leq |X| \quad \text{and} \quad E[|X|] < \infty.$$

Thus, by the DCT we get that when $n \to \infty$, $E[Z_n] \to E[Z] = 0$ and, since,

$$E[Z_n] = \int_\Omega Z_n dP = \int_\Omega (XI_{A_n})dP = \int_{A_n} XdP,$$

we are done.
In proving the second statement of the problem, one should notice that if A_n is defined as

$$A_n = \{\omega \in \Omega : |Y|(\omega) > n\},$$

then, as $E[|Y|] < \infty$, it must be $A_n \downarrow \emptyset$ as $n \to \infty$. From this fact it follows from the first part of the problem that

$$\int_{A_n} |Y|dP \to 0$$

as $n \to \infty$. In addition, if this is the case, it is always possible to find a positive integer c such that

$$\int_{A_n} |Y|dP < \epsilon, \forall n > c.$$

In addition, for every $B_n \in \mathcal{F}$, we have

$$\begin{aligned}\int_{B_n} |Y|dP &= \int_{B_n \cap \{|Y|>c\}} |Y|dP + \int_{B_n \cap \{|Y| \le c\}} |Y|dP \\ &\le \int_{\{|Y|>c\}} |Y|dP + c\int_{B_n} dP \\ &= \int_{\{|Y|>c\}} |Y|dP + cP(B_n) \le \epsilon + cP(B_n).\end{aligned}$$

Since $P(B_n) \to 0$ as $n \to \infty$, there exists a positive integer d such that $P(B_n) < \epsilon/c \ \forall n > d$. This proves that $\forall n > \max\{c, d\}$, it must be

$$\int_{B_n} |Y|dP \le \epsilon + c\frac{\epsilon}{c} = 2\epsilon$$

and, since ϵ is arbitrary, the proof of the second statement is complete.

PROBLEM 3.2.2. [3] On $((0,1), \mathcal{B}(0,1), \lambda)$ consider the following sequence of random variables:

$$\{X_n = n^\alpha I_{(1/(n+1),1/n)}(\omega)\}$$

where α is any number in $(1,2)$. Prove that $E[X_n] \to 0$ as $n \to \infty$.

[3] This problem is problem 3.9 in R. Durrett's Probability: Theory and Examples, 2^{nd} ed., 1996.

SOLUTION. Since $X_n \stackrel{a.s.}{\to} 0$, one is tempted to use the DCT and argue that $E[X_n] \to 0$ as $n \to \infty$. Unfortunately, this approach does not work here. In fact, in order to be able to use the DCT, one needs to find a r.v. Y such that $|X_n| \leq Y\ \forall n = 1, 2, \ldots$ and $E[Y] < \infty$. This is not possible. Assume it were. Then, there is a r.v. $Y \geq |X_n|$ for any choice of n and, hence,

$$E[Y] \geq \sum_{n=1}^{\infty} n^{\alpha}\left(\frac{1}{n} - \frac{1}{n+1}\right) = \sum_{n=1}^{\infty} \frac{n^{\alpha-1}}{n+1} \geq \sum_{n=1}^{\infty} \frac{1}{n+1} = \infty$$

and, thus, $E[Y] = \infty$, which is a contradiction.

Despite this, there is a fact that can help in a situation like the one above, where a sequence of r.v.'s is converging almost surely to some limit but not fast enough to use the DCT.

Proposition. *Suppose $X_n \stackrel{a.s.}{\to} X$ and there are continuous functions, $g(\cdot), h(\cdot) \geq 0$ with $g(x) \geq 0$ and $|h(x)|/g(x) \to 0$ as $|x| \to \infty$ and $E[g(X_n)] \leq k < \infty$ forall n. Then, $E[h(X_n)] \to E[h(X)]$.*

Proof. See R. Durrett, Probability: Theory and Examples, 2^{nd} ed. Thm 3.8. □

Thus, if we let $h(x) = x$, all we need is a function $g(x)$ that satisfies the assumptions of the Proposition above. If one tries functions like $g(x) \equiv |x|^{k/\alpha}$, it is easily seen that this function is easy to handle and that it must be $k = 2$ or $E[g(X_n)] = \infty$. Thus, let $g(x) = |x|^{2/\alpha}$. Then,

1. $g(x) \geq 0, \forall x$.
2. $\frac{|h(x)|}{g(x)} = |x|^{1-2/\alpha} \to 0$ as $|x| \to \infty$ since $\alpha \in (1, 2)$.
3. $E[g(X_n)] = E[|X_n|^{2/\alpha}] = E[n^{2/\alpha} I_{(1/(n+1), 1/n)}] = \frac{n}{n+1} < 1 < \infty$ forall n.

Thus, the Proposition above guarantees that

$$E[h(X_n)] \to E[h(x)]$$

that is

$$E[X_n] \to E[X] = E[0] = 0$$

as we were supposed to prove.

PROBLEM 3.2.3. Suppose that $f_n(x)$ and $g(x)$ are density functions such that $f_n(x) \to g(x)\ \forall x \in X$, the domain of definition. Show that

$$\lim_{n\to\infty} \int_X |f_n(x) - g(x)| dx = 0.$$

Show also that if X_n has density $f_n(x)$, X has density $g(x)$ and if $f_n(x) \to g(x)$ $\forall x \in X$, then

$$\lim_{n\to\infty} \sup_{A \in \sigma(X)} |P[X_n \in A] - P[X \in A]| = 0.$$

SOLUTION. For the first part of the problem, one should recall the following fact: $|x-y|= x-y+2\max\{y-x,0\}\ \forall x,y \in \mathbf{R}$.
Thus, it is possible to write

$$|f_n(x)-g(x)|= f_n(x)-g(x)+2\max\{g(x)-f_n(x),0\}$$

and

$$\begin{aligned}&\lim_{n\to\infty}\left[\int |f_n(x)-g(x)|dx\right]\\ &\quad=\lim_{n\to\infty}\left[\int f_n(x)-g(x)dx+2\int \max\{g(x)-f_n(x),0\}dx\right]\\ &\quad=\lim_{n\to\infty}\left[\int f_n(x)dx-\int g(x)dx+2\int \max\{g(x)-f_n(x),0\}dx\right]\end{aligned}$$

Now, $\int f_n(x)dx = 1$ as $f_n(\cdot)$ is a density function and the same is true of $\int g(x)dx$, therefore, we have

$$lim_{n\to\infty}\left[\int |f_n(x)-g(x)|dx\right]=\lim_{n\to\infty}\left[2\int \max\{g(x)-f_n(x),0\}dx\right].$$

As $\max\{g(x)-f_n(x),0\}\leq g(x)$ and $\int g(x)dx=1<\infty$ we can use the DCT for functions (see e.g., R. M. Dudley, Real Analysis and Probability, 1989; p. 101) and write

$$\begin{aligned}&\lim_{n\to\infty}\left[\int |f_n(x)-g(x)|dx\right]\\ &\quad=2\left[\int_H g(x)-\lim_{n\to\infty} f_n(x)dx+\int_{H^c}\lim_{n\to\infty} f_n(x)-g(x)dx\right]=0\end{aligned}$$

where $H\equiv\{x : x\in\mathbf{R}, f_n(x)\geq g(x)\}$.
The second part of the problem follows easily from the previous one and the fact that the sequence $\{X_n, n\geq 1\}$ and X have densities $f_n(\cdot)$ and $g(\cdot)$, respectively. Thus,

$$\begin{aligned}&\sup_{A\in\sigma(X)}|P[X_n\in A]-P[X\in A]|\\ &\quad=\sup_{A\in\sigma(X)}|\int_A f_n(x)-g(x)|dx\leq\sup_{A\in\sigma(X)}\int_A |f_n(x)-g(x)|dx\\ &\quad=\int_{\mathbf{R}}|f_n(x)-g(x)|dx\to 0\end{aligned}$$

as $n\to\infty$.

PROBLEM 3.2.4. Let $\{X_n, n\geq 1\}$ be a sequence of r.v.'s on the same probability space $(\Omega,\mathcal{F},P)$. We are interested in conditions which guarantee that

$$E[\sum_{n=1}^{\infty} X_n]=\sum_{n=1}^{\infty} E[X_n].$$

For each of the conditions listed below verify whether they suffice for the previous equality to hold or provide a counterexample.

i. $X_n \geq 0$ for all $n = 1, 2, \ldots$;

ii. $E[\sum_{n=1}^{\infty}|X_n|] < \infty$;

iii. $\sum_{n=1}^{\infty}|X_n| < \infty$ a.s.

SOLUTION. Condition i. is enough and this fact was already proved in Problem 3.1.1.
Condition ii. is also sufficient. This follows from an application of the DCT. Nonetheless, we have to be a little bit careful. In fact, let $Y_n = \sum_{i=1}^{n}|X_i|$ and $Y = \sum_{n=1}^{\infty}|X_n|$. Then, it is easy to see that we have $E[Y_n] \leq E[Y] = E[\sum_{n=1}^{\infty}|X_n|] < \infty$ for all $n = 1, 2, \ldots$ But, the DCT requires[4] also that $Y_n \overset{a.s.}{\to} Y$ as $n \to \infty$. We will show that this follows from the assumption $E[\sum_{n=1}^{\infty}|X_n|] < \infty$. To see why it is so, let $Z_n(\omega) = \sum_{i=1}^{n}|X_i(\omega)|$, $n \geq 1$. It is easily seen that, by the MCT, we must have

$$E[Z] = \sum_{i=1}^{\infty} E[|X_i|] < \infty,$$

where $Z(\omega) = \sum_{i=1}^{\infty} E[|X_i(\omega)|]$. This implies that $\sum_{i=1}^{\infty}|X_i| < \infty$ a.s. or, otherwise, $E[Z] = \infty$. In addition, we have also found that it is $\sum_{i=1}^{\infty} X_i < \infty$ a.s.. But, if the infinite sum is going to converge almost surely, we need $X_n \overset{a.s.}{\to} 0$ as $n \to \infty$. This last fact is incompatible with having $\liminf_n X_n(\omega) \neq \limsup_n X_n(\omega)$ (i.e. the limit of the sequence does not exist) for $\omega \in A \in \mathcal{F}$ and $P(A) > 0$, as in this case we do not have $X_n \overset{a.s.}{\to} 0$.
The third condition, on the contrary, does not imply the equality we would like to establish. In fact, let

$$X_n(\omega) = \begin{cases} n^2 I_{[0,1/n^2]}(\omega) & \text{if } n = 1, 3, 5, \ldots; \\ -X_{n-1}(\omega) & \text{if } n = 2, 4, 6, \ldots;. \end{cases}$$

[4] Actually, there is a more general version of the DCT than that usually found in textbooks. Here it is:

Theorem. *Let $\{X_n, n \geq 1\}$ be a sequence of $\bar{\mathbb{R}}$−valued r.v.'s defined on a probability space $(\Omega, \mathcal{F}, P)$ and let Y be another nonnegative r.v.'s defined on the same probability space and such that $|X_n| \leq Y$ a.s. and $E[Y] < \infty$. Then,*

$$-\infty < E[\liminf_n X_n] \leq \liminf_n E[X_n] \leq \limsup_n E[X_n] \leq E[\limsup_n X_n] < \infty.$$

In addition, if $X = \lim_n X_n$ exists a.s. we can also write

$$E[|X|] < \infty \text{ and } \lim_n E[X_n] = E[X].$$

Proof. Apply Fatou's Lemma to the sequences of random variables $\{Y - X_n, n \geq 1\}$ and $\{Y + X_n, n \geq 1\}$. □

Now, for any $\epsilon > 0$, we have that $P[|[|X_n] > \epsilon] = 1/n^2]$. Then, a simple application of the Borel Lemma gives that $P[\{\omega : X_n(\omega) > 0 \text{ i.o}\}] = 0$. This is equivalent to stating that $\sum_{n=1}^{\infty} |X_n| < \infty$ a.s. and, thus, this sequence of r.v.'s satisfies condition iii.. Nonetheless, it is easy to check that

$$\sum_{i=1}^{n} X_i = \begin{cases} X_n & \text{if } n = 1,3,5,\ldots; \\ 0 & \text{if } n = 2,4,6,\ldots;. \end{cases}$$

and, since $E[X_i] = 1$ for all $i \geq 1$, we have that $E[\sum_{i=1}^{n} X_i] = 0$ or 1 depending on whether n is odd or even, respectively. Hence, $E[\sum_{n=1}^{\infty} X_n]$ does not exist but, $\sum_{n=1}^{\infty} E[X_n] = \infty$.

3.3 Uniform Integrability

PROBLEM 3.3.1. Let $X_1, X_2, \ldots$ be a sequence of $\mathbf{R}$-valued random variables defined on a common probability space $(\Omega, \mathcal{F}, P)$. Under each of the following assumptions, either prove that $\lim_n E[|X_n|] = 0$ or provide a counterexample:

(a) $X_n \to 0$ uniformly as $n \to \infty$;

(b) $E[X_n] \to 0$ as $n \to \infty$;

(c) $X_1 \geq X_2 \geq X_3, \geq \ldots$ and $X_1 \sim \exp(1)$;

(d) $\sup_n \text{Var}(X_n) < \infty$.

SOLUTION. Since, by assumption, $X_n \to 0$ uniformly a.s. this implies that for any $\epsilon > 0$ there exists N_ϵ such that

$$|X_n(\omega)| \leq \epsilon \ \forall n > N_\epsilon, \forall \omega \in A \subseteq \Omega$$

and $P(A^c) = 0$. This implies that

$$E[|X_n|] \leq \epsilon \cdot P(A) + kP(A^c) = \epsilon$$

as the X_n's are $\mathbf{R}$-valued r.v.'s and hence $k < \infty$. Finally, as ϵ is arbitrary, it follows that assumption (a) suffices to have $\lim_m E[|X_n|] = 0$.

Assumption (b) is not enough to guarantee that $\lim_n E[|X_n|] = 0$. A counterexample is the following. Let Y be a Uniform[0, 1] r.v., and define

$$X_n = \begin{cases} -n^2 & \text{if } Y \in [0, 1/n^2]; \\ 0 & \text{if } Y \in (1/n^2, (n^2-1)/n^2); \\ n^2 & \text{if } Y \in [(n^2-1)/n^2, 1]. \end{cases}$$

It is easy to prove that $X_n \stackrel{a.s.}{\to} 0$ when $n \to \infty$ (use the Borel-Cantelli Lemma), $E[X_n] = 0$ for all $n \geq 1$ and, thus, $\lim_n E[X_n] = 0$. But, $E[|X_n|] = 2$ for all $n \geq 1$ and, therefore $\lim_n E[|X_n|] \neq 0$.

On the contrary, assumption (c) suffices to obtain $\lim_n E[|X_n|] = 0$. In fact, we note first that $X_i \geq 0$ for all $i \geq 1$. X_1 is positive because, by assumption, $X_1 \sim$ Exponential(1), and the other X_i's cannot be negative because of the combination of the two assumptions $X_n \xrightarrow{a.s.} 0$ and $X_n \geq X_{n+1}$, for all $n \geq 1$. To see that this is the case, assume that there is some $i \geq 2$ such that

$$X_i(\omega) < 0 \;\forall \omega \in A$$

and $P[A] > 0$. Because of the assumption $X_1 \geq X_2 \geq X_3 \geq \ldots$, this means that $X_j(\omega) < 0$ $\forall \omega \in A$ and $\forall j \geq i$. This means that there is a set A where $P[\lim_n |X_n| > \epsilon] \geq P[A] > 0$. Hence, $X_n \stackrel{a.s.}{\not\to} 0$ as $n \to \infty$ and this is a contradiction.

Since all X_n's are nonnegative, we have that $|X_n| = X_n \leq X_1$, $n \geq 2$, and $E[X_1] = 1 < \infty$. As by assumption, $X_n \xrightarrow{a.s.} 0$ as $n \to \infty$ we can use the DCT to write

$$\lim_n E[|X_n|] = \lim_n E[X_n] = E[0] = 0.$$

The last case is a little bit more complicated. Essentially, we want to be able to use the following important result.

Theorem. *Let $\{X_n, n \geq 1\}$ be a sequence of $\mathbf{R}-$valued r.v.'s whose limit X exists almost surely. Assume that $E[|X_n|] < \infty$ for all n. Then the following statements are equivalent:*

1. *the family $\{X_n, n \geq 1\}$ is u.i.;*
2. *$E[|X|] < \infty$, and $\lim_n E[|X_n - X|] = 0$;*
3. *$\lim_n E[|X_n|] = E[|X|] < \infty$.*

Each of these conditions implies

4. *$\lim_n E[X_n] = E[X]$.*

Proof. See B. Fristedt and L. Gray, A Modern Approach to Probability Theory, 1997, Theorem 12 page 108. □

Thus, if we can prove that the sequence of random variables is u.i. and that $E[|X_n|] < \infty$, we are done because we can use the third statement of the theorem above. In order to do this we need two lemmas.

Lemma 1. *If $X_n \xrightarrow{a.s.} 0$ as $n \to \infty$ and, in addition, $\sup_n Var[X_n] < \infty$, then $E[|X_n|] \not\to \infty$.*

Proof. This is the consequence of the Chebyshev's Inequality. In fact, we can always write that

$$P[|X_n - E[X_n]|] < k(Var[X_n])^{1/2}] \geq 1 - k^{-2}$$

or

$$P[E[X_n] - k(Var[X_n])^{1/2} < X_n < E[X_n] + k(Var[X_n])^{1/2}] \geq 1 - k^{-2}.$$

Assume now that $E[X_n] \to \pm\infty$. Because of the assumption that $\sup_n Var[X_n] < \infty$, this implies that for $k \geq 1$ and for any $\epsilon > 0$, it must be

$$P[\lim_n |X_n| > \epsilon] > 0$$

which contradicts the assumption that $X_n \xrightarrow{a.s.} 0$ as $n \to \infty$. □

Now, by assumption, it is $\sup_n Var[X_n] < \infty$. As we just proved that $\sup_n E[X_n] < \infty$, the formula for the variance tells us that this implies also $\sup_n E[X_n^2] < \infty$ and this, in turn, gives that $E[|X_n|] < \infty$. We are left with having to establish that the assigned sequence of r.v.'s is u.i. This is a consequence of

Lemma 2. *Let $\{X_n, n \geq 1\}$ be a family of $\bar{\mathbf{R}}$−valued r.v.'s on some probability space $(\Omega, \mathcal{F}, P)$ and suppose that there exists $p > 1$ and $k < \infty$ such that $E[|X_n|]^p] \leq k$, $\forall n \geq 1$. Then, the family $\{X_n, n \geq 1\}$ is uniformly integrable.*

Proof. Let A_c be the event $\{\omega : |X_n(\omega)| \geq c\}$ and let I_{A_c} be its indicator function. Then, one can write the following inequality

$$E[|X_n|I_{A_c}] = E[|X_n|^{1-p}|X_n|^p I_{A_c}] \leq c^{1-p}E[|X_n|^p] \leq \frac{k}{c^{p-1}}.$$

Now, since $p > 1$ we have found that

$$\lim_{c\to\infty} \sup_n E[|X_n|I_{A_c}] = 0.$$

□

We have already established that in our case it is $\sup_n E[X_n^2] < \infty$ and, by lemma 2, this suffices to state that the sequence of r.v.'s is u.i.. Finally, using part (3) of the theorem, gives

$$\lim_n E[|X_n|] = E[0] = 0.$$

PROBLEM 3.3.2. Let $\{X_n, n \geq 1\}$ be a sequence of r.v.'s defined on the same probability space $(\Omega, \mathcal{F}, P)$. Prove that if there exists a random variable Y on the same probability space such that $|X_n| \leq Y$ for any $n = 1, 2, \ldots$ and $E[Y] < \infty$, then the sequence $\{X_n, n \geq 1\}$ is uniformly integrable.
Prove also that this condition is only sufficient but not necessary

SOLUTION. To show that the existence of a r.v. Y such that $|X_n| \leq Y$, $n = 1, 2, \ldots$ and $E[Y] < \infty$ implies u.i., it suffices to notice that

$$E[|X_n|] \leq E[Y],\ \forall n \geq 1 \implies \sup_n E[|X_n|] \leq E[Y] < \infty$$

and, in addition, since $E[Y] < \infty$ this means that $\forall \epsilon > 0$, $\exists \delta_\epsilon$ and $A \in \mathcal{F}$ such that $P(A) < \delta_\epsilon$ and $\int_A Y dP < \epsilon$. But, if this is the case, it must be also that $\int_A |X_n| dP < \epsilon$. The next theorem gives the desired conclusion.

Theorem. *A sequence of r.v.'s $\{X_n, n \geq 1\}$ on a probability space $(\Omega, \mathcal{F}, P)$ is uniformly integrable iff*

1. $\sup_n E[|X_n|] < \infty$;
2. $\forall \epsilon > 0\ \exists \delta > 0$ *such that* $\int_E |X_n| dP < \epsilon\ \forall n \geq 1$, *whenever* $E \in \mathcal{F}$ *and* $P[E] < \delta$.

To prove that the condition is only sufficient, it is enough to provide a counterexample. To this purpose, consider the following probability space $\mathbf{N}, 2^{\mathbf{N}}, P)$ where the probability measure P is defined as follows

$$P[n] = \frac{\alpha}{n(\log n)^p} \quad p > 1; n \in \mathbf{N}$$

where $\alpha = (\sum_{n=1}^{\infty} n(\log n)^p)^{-1}$ is a normalizing constant. On this probability space, define the sequence of r.v.'s $\{X_n, n \geq 1\}$ by letting

$$X_n(\omega) = n \cdot I_{\{n\}}(\omega).$$

Now,

$$\int_{|X_n| > c} = \begin{cases} 0 & \text{if } n \leq c; \\ \frac{\alpha}{(\log n)^p} & \text{if } n > c. \end{cases}$$

In either case, we have

$$\int_{|X_n| > c} \leq \frac{\alpha}{(\log c)^p} \to 0$$

as $c \to \infty$. This proves that the sequence of random variables we just defined is uniformly integrable. Now, assume that there exists a random variable Y on the same probability space which dominates the sequence. This implies that

$$Y(\omega) \geq n \cdot I_{\{n\}}(\omega)$$

and, thus,

$$E[Y] = \sum_{n=1}^{\infty} n \cdot P[\{n\}] = \alpha \sum_{n=1}^{\infty} \frac{1}{(\log n)^p} = \infty.$$

PROBLEM 3.3.3. Let $\{X_n, n \geq 1\}$ be a sequence of r.v.'s on some probability space $(\Omega, \mathcal{F}, P)$. If $\{X_n, n \geq 1\}$ is dominated by some random variable Y on the same probability space and $E[|Y|] < \infty$, then $\{X_n, n \geq 1\}$ is uniformly integrable.

SOLUTION. To prove that $\{X_n, n \geq 1\}$ is uniformly integrable we need to show that $\lim_{c\to\infty} \int_{|X_n|>c} |X_n| dP = 0$. To this purpose, using the fact that $\{X_n, n \geq 1\}$ is dominated by Y, it is possible to write

$$\lim_{c\to\infty} \int_{|X_n|>c} |X_n| dP \leq \lim_{c\to\infty} \int_{|X_n|>c} Y dP \leq \lim_{c\to\infty} \int_{Y>c} Y dP$$

where the last inequality follows from the fact that $\{\omega : |X_n|(\omega) > c\} \subseteq \{\omega : Y(\omega) > c\}$. Now, to show that the last term goes to 0 as $c \to \infty$, we can use the following argument. Let $\Lambda_c = \{\omega : Y(\omega) > c\}$, then we clearly have that $\Lambda \downarrow \emptyset$ or, equivalently, $\Lambda_c^c \uparrow \Omega$. This means that $I_{\Lambda_c^c} Y \uparrow Y$ as $c \to \infty$ and an application of the MCT gives $E[I_{\Lambda_c^c} Y] \to E[Y]$ as $c \to \infty$. Therefore, we have established that

$$\lim_{c\to\infty} \int_{Y>c} Y dP = \int_{\Omega} Y dP - \lim c \to \infty \int_{Y \leq c} Y dP = 0.$$

PROBLEM 3.3.4. Let $\{X_n; n \geq 1\}$ be a sequence of u.i. random variables defined on some probability space $(\Omega, \mathcal{F}, P)$. Define $Y_n = \frac{1}{n} \sum_{i=1}^{n} X_i$, $n = 1, 2, \ldots$ Prove that $\{Y_n; n \geq 1\}$ is also a u.i. sequence of random variables on $(\Omega, \mathcal{F}, P)$.

SOLUTION. Instead of using the original criterion to prove u.i. we give an alternative criterion whose proof can be found in most textbooks in probability theory (e.g. K. L. Chung, A course in probability theory, 2nd ed. 1974; thm 4.5.3)

Theorem. *A sequence of r.v.'s $\{X_n, n \geq 1\}$ on a probability space $(\Omega, \mathcal{F}, P)$ is uniformly integrable iff*

1. $\sup_n E[|X_n|] < \infty$;
2. $\forall \epsilon > 0$ $\exists \delta > 0$ *such that* $\int_E |X_n| dP < \epsilon$ $\forall n \geq 1$, *whenever* $E \in \mathcal{F}$ *and* $P[E] < \delta$.

Thus, since, by assumption $\{X_n, n \geq 1\}$ is u.i., it satisfies the two conditions of the theorem just stated. Then, we have

$$E[|Y_n|] \leq \frac{1}{n} \sum_{i=1}^{n} E[|X_i|] \leq \sup_n E[|X_n|] \equiv M < \infty.$$

Since this inequality holds for every $n = 1, 2, \ldots$, we have also

$$\sup_n E[|Y_n|] < \infty.$$

In addition, let E and δ be the same as in the statement of the theorem. It turns out that for all $n \geq 1$

$$\int_E |Y_n| dP \leq \int_E \frac{1}{n}\sum_{i=1}^{n}|X_i| dP = \frac{1}{n}\sum_{i=1}^{n}\int_E |X_i| dP < \frac{\epsilon n}{n} = \epsilon.$$

As we have shown that $\{Y_n, n \geq 1\}$ satisfies 1. and 2. of the theorem, it follows that the sequence of r.v.'s is u.i.

PROBLEM 3.3.5. Let $\{X_n, n \geq 1\}$ and $\{Y_n, n \geq 1\}$ be u.i. sequences of r.v.'s on some probability space $(\Omega, \mathcal{F}, P)$. Prove that the sequence $\{X_n + Y_n, n \geq 1\}$ is also u.i..

SOLUTION. Again we will use the following

Theorem. *A sequence of r.v.'s $\{X_n, n \geq 1\}$ on a probability space $(\Omega, \mathcal{F}, P)$ is uniformly integrable iff*

1. $\sup_n E[|X_n|] < \infty$;
2. $\forall \epsilon > 0\ \exists \delta > 0$ *such that* $\int_E |X_n| dP < \epsilon\ \forall n \geq 1$, *whenever* $E \in \mathcal{F}$ *and* $P[E] < \delta$.

Now,

$$\sup_n E[|X_n + Y_n|] \leq \sup_n \{E[|X_n|] + E[|Y_n|]\} \leq \sup_n E[|X_n|] + \sup_n E[|Y_n|] < \infty$$

as, by assumption, the original sequences of r.v.'s are u.i.. This takes care of the first condition. To verify the second condition we use again the fact that $\{X_n, n \geq 1\}$ and $\{Y_n, n \geq 1\}$ are u.i., hence, $\forall \epsilon > 0$ we can find two numbers δ_ϵ and δ'_ϵ such that $\forall E, E' \in \mathcal{F}$ with $P[E] < \delta_\epsilon$, and $P[E'] < \delta'_\epsilon$, we have

$$\int_E |X_n| dP < \epsilon, \text{ and } \int_{E'} |Y_n| dP < \epsilon,$$

respectively. Let $F = E \cap E'$. Clearly, it is $F \in \mathcal{F}$ and $P[F] < \min\{\delta_\epsilon, \delta'_\epsilon\} \equiv \delta_\eta$. In addition, we have

$$\begin{aligned}\int_F |X_n + Y_n| dP &\leq \int_F |X_n| dP + \int_F |Y_n| dP \\ &\leq \int_E |X_n| dP + \int_{E'} |Y_n| dP < 2\epsilon \equiv \eta.\end{aligned}$$

Since ϵ is an arbitrary positive number, so is η. Thus, we have proved that $\forall \eta > 0$, $\exists \delta_\eta > 0$ such that $\forall F \in \mathcal{F} : P[F] < \delta_\eta$, we have

$$\int_F |X_n + Y_n| dP < \eta.$$

This verifies the second condition of the theorem above and proves that the sequence of r.v.'s $\{X_n + Y_n, n \geq 1\}$ is u.i. as we were supposed to show.

3.4 Borel-Cantelli Lemmas

PROBLEM 3.4.1. Let $\{X_n, n \geq 1\}$ be a sequence of identically distributed r.v.'s on a probability space $(\Omega, \mathcal{F}, P)$. Show

1. $X_n/n \xrightarrow{P} 0$ as $n \to \infty$;
2. if $E[|X_1|] < \infty$, then $X_n/n \xrightarrow{a.s.} 0$ as $n \to \infty$;
3. if $X_1, X_2, \ldots$ are independent, then $X_n/n \xrightarrow{a.s.} 0$ implies $E[|X_1|] < \infty$.

Conclude further that if $X_1, X_2, \ldots$ are i.i.d. and $E[|X_1|] = \infty$, then $X_n/n \xrightarrow{P} 0$ but $X_n/n \not\xrightarrow{a.s.} 0$ as $n \to \infty$.

SOLUTION. The first conclusion follows from the following argument. For all $\epsilon > 0$,

$$P[|X_n|/n > \epsilon] = P[|X_n| > n\epsilon] = F(-n\epsilon) + 1 - F(n\epsilon).$$

Thus,

$$\lim_n P[|X_n|/n > \epsilon] = \lim_n [F(-n\epsilon) + 1 - F(n\epsilon)] = 0.$$

The second statement can be shown to be correct using the inequality

$$E[|X_1|] = \int_0^\infty P[|X_1| > x]dx \geq \sum_{n=1}^\infty P[|X_n| > n\delta]$$

where δ is any positive number less than 1. Now, if $E[|X_1|] < \infty$, we have also $\sum_{n=1}^\infty P[|X_n| > n\delta] < \infty$. Then, the Borel Lemma and the arbitrariness of δ imply that $P[|X_n|/n > \delta \text{ i.o.}] = 0$ and, therefore, $X_n/n \xrightarrow{a.s.} 0$ as $n \to \infty$.
Assume that $E[|X_1|] = \infty$. This implies that

$$\infty = E[|X_1|] = \int_0^\infty P[|X_1| > x]dx \leq \sum_{n=0}^\infty P[|X_n| > n].$$

But, since the X_n's are assumed to be independent, the Borel-Cantelli Lemma tells us that this is the same as

$$P[|X_n|/n > 1 \text{ i.o.}] = 1$$

and $|X_n|/n \not\xrightarrow{a.s.} 0$ as $n \to \infty$. This contradicts the original assumption and, hence, it must be $E[|X_1|] < \infty$.

For an example of a sequence of i.i.d. random variables $\{X_n, n \geq 1\}$ such that $E[X_1] = \infty$, $X_n/n \xrightarrow{P} 0$, and $X_n/n \stackrel{a.s.}{\not\to} 0$ as $n \to \infty$ one can look at the following sequence of i.i.d random variables defined by

$$X_1 = 2^m \text{ with probability } 2^{-m}; m \geq 1.$$

For this sequence of i.i.d. r.v.'s we have

$$E[|X_1|] = E[X_1] = \sum_{m=1}^{\infty} 2^m \cdot 2^{-m} = \sum_{m=1}^{\infty} 1 = \infty,$$

$X_n/n \xrightarrow{P} 0$ as $n \to \infty$, since in part 1 the conclusion was independent from the assumptions of infinite expectation and independence, but $X_n/n \stackrel{a.s.}{\not\to} 0$ as $n \to \infty$. This is shown below. Letting $\epsilon = 1/2^k$, one realizes that the condition

$$P[|X_1| > \frac{n}{2^k}] = \frac{1}{2^m}$$

for $2^{m+k} \leq n < 2^{m+k+1}$. In fact, for $n \in [2^{m+k}, 2^{m+k-1})$, we have

$$P[|X_1| > \frac{n}{2^k}] = P[|X_1| > 2^m] = \sum_{i=1}^{\infty} \frac{1}{2^{m+i}} = \frac{1}{2^{m+1}} \sum_{i=0}^{\infty} \frac{1}{2^i} = \frac{1}{2^m}.$$

Then, always assuming $\epsilon = 2^{-k}$, and using the fact that by the definition of X_1 it is $P[|X_1| > n/2^k] = 1$ for $1 \leq n \leq 2^{k+1} - 1$, we can write

$$\begin{aligned}
\sum_{n=1}^{\infty} P[|X_1| > \frac{n}{2^k}] &= 2^{k+1} - 1 + \sum_{n=2^{k+1}}^{\infty} P[|X_1| > n/2^k] \\
&= \sum_{m=1}^{\infty} P[|X_1| > 2^{m+k}] \times \ \#\text{integers in}[2^{m+k}, 2^{m+k+1} - 1] \\
&= \sum_{m=1}^{\infty} \frac{1}{2^m} \times 2^{m+k} = \sum_{m=1}^{\infty} 2^k = \infty.
\end{aligned}$$

Thus, as for the particular sequence of random variables defined above the Borel Cantelli condition fails, we can conclude that $X_n/n \stackrel{a.s.}{\not\to} 0$ as $n \to \infty$.

PROBLEM 3.4.2. Prove that if $\{X_n, n \geq 1\}$ is a sequence of i.i.d. r.v.'s such that $E[|X_1|] = \infty$ and $S_n = \sum_{i=1}^{n} X_i$, then

$$P[\lim_n \frac{S_n}{n} \text{ exists and is finite}] = 0.$$

SOLUTION. From the previous problem we know that

$$P[|X_n| > n \text{ i.o}] = 1.$$

In addition, it is possible to write

$$\frac{S_n}{n} - \frac{S_{n+1}}{n+1} = \frac{S_n}{n(n+1)} - \frac{X_{n+1}}{n}.$$

Letting A be the event $\{\omega \in \Omega : \lim_n \frac{S_n}{n}$ exists and is finite$\}$, it follows easily that if $\omega \in A$, it must be

$$\frac{S_n(\omega)}{n(n+1)} \to 0$$

as $n \to \infty$. Combining together the statements elaborated so far, we have that for any $\omega \in A \cap \{\omega : |X_n(\omega)| > n \text{ i.o.}\}$

$$|\frac{S_n(\omega)}{n} - \frac{S_{n+1}(\omega)}{n+1}| > \alpha \text{ i.o.}$$

where α is any positive number less than 1. This implies that $\frac{S_n}{n}$ does not converge to a finite limit with probability 1. Hence, $\omega \notin A$ and this is a contradiction. We have, thus, showed that it must be

$$A \subseteq \{\omega : |X_n(\omega)| > n \text{ i.o.}\}^C$$

which, together with the fact that $P[\{\omega : |X_n(\omega)| > n \text{ i.o.}\}] = 1$, implies $P[A] = 0$.

PROBLEM 3.4.3. Let $A_1, A_2, \ldots,$ be a sequence of events in a probability space $(\Omega, \mathcal{F}, P)$ and set $A = \limsup_n A_n$. If $\sum_{n=1}^{\infty} P(A_n) = \infty$, and

$$\limsup_{n\to\infty} \frac{(\sum_{i=1}^{n} P(A_n))^2}{\sum_{i=1}^{n}\sum_{j=1}^{n} P(A_i \cap A_j)} = \alpha > 0,$$

then $P[A] \geq \alpha$. Show that if the events are negatively correlated, then $P[A] = 1$. This is the **Kochen-Stone Lemma** which extends the Borel-Cantelli Lemma to sequences of non independent random variables. Hence, the Borel-Cantelli Lemma holds not only for sequences of independent r.v.'s but for sequences of negatively correlated or noncorrelated r.v.'s as well.

SOLUTION.[5] To prove the Kochen-Stone Lemma, we need two lemmas:

Lemma 1. *Let $Y \geq 0$, $E[Y^2] < \infty$, and $E[Y] > a$. Then,*

$$P[Y > a] \geq \frac{(E[Y-a])^2}{E[Y^2]}.$$

[5] There is more than one way to prove this result. The approach used here is that suggested by Durrett (see R. Durrett, Probability: Theory and Examples, 2^{nd} ed. Problem 6.20, page 55). For a slightly different solution see B. Fristedt and L. Gray, A Modern Approach to Probability Theory, 1997 pp. 78-9.

Proof. By applying the Cauchy-Schwarz Inequality, one gets

$$E[YI_{[Y>a]}] \leq (E[Y^2])^{1/2} \cdot (E[I^2_{[Y>a]}])^{1/2}$$

or, equivalently,

$$(E[YI_{[Y>a]}])^2 \leq E[Y^2] \cdot E[I^2_{[Y>a]}].$$

Now, $E[I^2_{[Y>a]}] = P[Y > a]$ and, since

$$Y \cdot I_{[Y>a]} = \begin{cases} 0 & \text{if } Y \leq a; \\ Y & \text{if } Y > a \end{cases}$$

it follows that

$$Y \cdot I_{[Y>a]} \geq Y - a \ a.s.$$

which, in turn, implies

$$E[Y \cdot I_{[Y>a]}] \geq E[Y - a] = E[Y] - a.$$

This proves our statement

$$(E[Y] - a)^2 \leq E[Y^2]P[Y > a].$$

□

Lemma 2. *Let $A_1, A_2, \ldots,$ be events in a probability space $(\Omega, \mathcal{F}, P)$. Then,*

$$P[\limsup_{n\to\infty} A_n \geq \limsup_{n\to\infty} P(A_n) \geq \liminf_{n\to\infty} P(A_n) \geq P(\liminf_{n\to\infty} A_n).$$

Proof. Let $B_n = \cap_{k=n}^{\infty} A_n$. Then, $\{B_n, n \geq 1\}$ is an increasing sequence of events such that $\lim_{n\to\infty} B_n = \liminf_{n\to\infty} A_n$. Hence, by the Continuity of Measure Theorem, we have

$$\lim_{n\to\infty} P[B_n] = P[\liminf_{n\to\infty} A_n].$$

On the other hand, $P[B_n] \leq P[A_j], \forall j \geq n$; thus, it is possible to write

$$P[B_n] \leq \inf\{P[A_j] : j \geq n\}$$

and letting $n \to \infty$, we have

$$P[\liminf_{n\to\infty}] = \lim_{n\to\infty} P[B_n] \leq \liminf_{n\to\infty} P[A_n].$$

Using a similar argument one can show that

$$P[\limsup_{n\to\infty} A_n] \geq \limsup_{n\to\infty} P[A_n].$$

□

Let $Y_n = I_{A_n}$ be the indicator function of the event A_n. In addition, set $X_n = \sum_{k=1}^n Y_k$ and $Z_n = X_n/E[X_n]$. It is easily seen that $E[Z_n] = 1$ and

$$E[Z_n^2] = E[Z_n Z_n] = \frac{E[\sum_{i=1}^n Y_i \sum_{j=1}^n Y_j]}{(E[X_n])^2} = \frac{\sum_{i=1}^n \sum_{j=1}^n P[Y_i Y_j]}{(\sum_{i=1}^n P(A_i))^2}.$$

Since Z_n as defined above is a r.v. that satisfies $E[Z_n] = 1$, $E[Z_n^2] < \infty$, for all $n = 1, 2, \ldots$, then for any $\eta < 1$, we can use lemma 1 and write

$$P[Z_n > \eta] \geq \frac{(E[Z_n] - \eta)^2}{E[Z_n^2]} = \frac{(1-\eta)^2}{E[Z_n^2]}.$$

The key fact to establish the first result is the following:

$$P[A] \equiv P[\limsup_n A_n] \geq P[\limsup_n Z_n > \eta].$$

To understand why this is so[6], assume that $\omega \notin A$. This means that $\lim_{n\to\infty} X_n(\omega)$ is a finite number and, thus, $\lim_{n\to\infty} Z_n(\omega) = 0$, since by assumption, $E[X_n] = \sum_{i=1}^n P[A_i] \to \infty$ as $n \to \infty$. This is the same as saying that for any $\eta > 0$, $Z_n(\omega) \leq \eta$ i.o.. Hence, if $\omega \in A$, it must satisfy the requirement $Z_n(\omega) > \eta$ i.o. or, that is to say, $\omega \in \{\omega : \limsup_{n\to\infty} Z_n(\omega) > \eta\}$. Now, using lemma 2, we can write

$$P[\limsup_{n\to\infty} Z_n > \eta] \geq \limsup_{n\to\infty} P[Z_n > \eta] \geq \frac{1-\eta^2}{\limsup_{n\to\infty} E[Z_n^2]} = (1-\eta^2)\alpha$$

using the second assumption of the problem. Since η is an arbitrary number in $(0,1)$, it follows that $P[A] \geq \alpha$ as we were supposed to show.

If we introduce the extra assumption that the events $A_1, A_2, \ldots$ are negatively correlated, this means that $P[A_i \cap A_j] \leq P[A_i]P[A_j]$, $\forall i, j = 1, 2, \ldots$ Then,

$$\begin{aligned}
\frac{(\sum_{i=1}^n P(A_i))^2}{\sum_{i=1}^n \sum_{j=1}^n P(A_i \cap A_j)} &= \frac{\sum_{n=1}^\infty P(A_n)^2 + \sum_{i=1}^n \sum_{j=1, j\neq i}^n P[A_i]P[A_j]}{\sum_{i=1}^n \sum_{j=1, j\neq i}^n P[A_i]P[A_j] + \sum_{i=1}^n P[A_i]^2} \\
&\geq \frac{\sum_{i=1}^n \sum_{j=1, j\neq i}^n P[A_i]P[A_j]}{\sum_{i=1}^n \sum_{j=1, j\neq i}^n P[A_i]P[A_j] + \sum_{i=1}^n P[A_i]} \\
&= \frac{1}{1 + \frac{\sum_{i=1}^n P[A_i]}{\sum_{i=1}^n \sum_{j=1, j\neq i}^n P[A_i]P[A_j]}} \\
&= \frac{1}{1 + \frac{\sum_{i=1}^n P[A_i]}{(\sum_{i=1}^n P[A_i])^2 - \sum_{i=1}^n P[A_i]^2}}
\end{aligned}$$

Now, since,

$$\frac{\sum_{i=1}^n P[A_i]}{(\sum_{i=1}^n P[A_i])^2 - \sum_{i=1}^n P[A_i]^2} = \frac{1}{\sum_{i=1}^n P[A_i] - \frac{\sum_{i=1}^n P[A_i]^2}{\sum_{i=1}^n P[A_i]}} \to 0$$

[6]Essentially, we prove that $A^c \subseteq \{\omega : \limsup_n Z_n(\omega) > \eta\}^c$. This gives $A \supseteq \{\omega : \limsup_n Z_n(\omega) > \eta\}$.

as $n \to \infty$ as, by assumption, $\sum_{i=1}^{\infty} P[A_i] = \infty$ we can use the statement just proved to write

$$P[A] \geq \limsup_{n\to\infty} \frac{(\sum_{i=1}^{n} P(A_n))^2}{\sum_{i=1}^{n}\sum_{j=1}^{n} P(A_i \cap A_j)} = 1.$$

Hence $P[A] = 1$.

PROBLEM 3.4.4. Let $\{X_n, n \geq 1\}$ be a sequence of r.v.'s defined on a common probability space $(\Omega, \mathcal{F}, P)$. Assume, in addition, that each random variable is uniformly distributed on $[0, 1]$. Prove that for all $\alpha > 1$

$$\lim_n \frac{1}{n^\alpha X_n} = 0 \ a.s.$$

SOLUTION. Let $\epsilon > 0$ be any arbitrary number, and set $A_n = \{\omega : \frac{1}{n^\alpha X_n(\omega)} > \epsilon\}$. It follows that

$$P[A_n] = P[\{\omega : \frac{1}{n^\alpha X_n(\omega)} > \epsilon\}] = \frac{1}{\epsilon n^\alpha}.$$

Hence, it is easily seen that

$$\sum_{n=1}^{\infty} P[A_n] = \sum_{n=1}^{\infty} \frac{1}{\epsilon n^\alpha} = \frac{1}{\epsilon}\sum_{n=1}^{\infty} \frac{1}{n^\alpha} < \infty,$$

using the assumption that $\alpha > 1$. Thus, by the Borel Lemma we have that $\forall \epsilon > 0$

$$P[\frac{1}{n^\alpha X_n(\omega)} > \epsilon \text{ i.o.}] = 0.$$

Since $\epsilon > 0$ is arbitrary, this is the same as saying that

$$\lim_n \frac{1}{n^\alpha X_n} = 0 \text{ a.s.}$$

as we were supposed to prove.

PROBLEM 3.4.5. Let $\{X_n, n \geq 1\}$ be a sequence of pairwise negatively correlated or uncorrelated r.v.'s each having Bernoulli distribution. Prove that

$$\sum_{n=1}^{\infty} X_n = \begin{cases} \infty & \text{a.s.} \\ < \infty & \text{a.s.} \end{cases}$$

iff

$$E[\sum_{n=1}^{\infty} X_n] = \begin{cases} \infty & \text{a.s.} \\ < \infty & \text{a.s.} \end{cases}$$

Provide an example that shows that the assumption about the correlation of the random variables in the sequence cannot be dropped.

SOLUTION. Since $X_n \sim$Bernoulli(p_n), for any $\epsilon > 0$, we have $P[|X_n|> \epsilon] = P[X_n = 1] = p_n$. Thus, if $\sum_{n=1}^{\infty} P[|X_n|> \epsilon] = \sum_{n=1}^{\infty} p_n = \infty$, we can use the Borel-Cantelli Lemma and conclude that $P[X_n = 1, \textit{ i.o.}] = 1$. This clearly implies that $\sum_{n=1}^{\infty} X_n = \infty$. In addition, we have $E[X_n] = p_n, n = 1, 2, \ldots$ and, since $X_n \geq 0, n = 1, 2, \ldots$, the MCT implies that $E[\sum_{n=1}^{\infty} X_n] = \sum_{n=1}^{\infty} E[X_n] = \sum_{n=1}^{\infty} p_n = \infty$. This proves one implication. On the contrary, if we assume that $\sum_{n=1}^{\infty} p_n < \infty$, the Borel Lemma implies that $P[X_n = 1 \textit{ i.o.}] = 0$. Hence, only finitely many of the X_n's are different from zero. This implies that $\sum_{n=1}^{\infty} X_n < \infty$.

To show that the assumption about the correlation cannot be dropped it suffices to look at the following example. Let $Y \sim$ Uniform[0, 1] r.v. and let

$$X_n(\omega) = \begin{cases} 1 & \text{if } Y(\omega) \in [0, 1/n]; \\ 0 & \text{otherwise.} \end{cases}$$

It is easily seen that the sequence of random variables just defined is positively correlated. Now, $E[X_n] = P[X_n = 1] = 1/n$ and this implies that $E[\sum_{n=1}^{\infty} X_n] = \infty$. But, at the same time, we have $P[\sum_{n=1}^{\infty} X_n = \infty] = P[Y = 0] = 0$ and, thus, $\sum_{n=1}^{\infty} X_n < \infty$ a.s..

PROBLEM 3.4.6. Prove that for any sequence of random variables $\{X_n, n \geq 1\}$ on some probability space $(\Omega, \mathcal{F}, P)$ there exists a sequence of constants $\{a_n, n \geq 1\}$ such that $X_n/a_n \xrightarrow{a.s.} 0$.

SOLUTION. Let's assume that the X_n's are nonnegative and, in addition, let's choose the a_n's in a way for which $a_n/n \to \infty$ as $n \to \infty$ and $P[X_n > a_n/n] < 2^{-n}$.[7] If we set $Y_n \equiv X_n/a_n$, $n \geq 1$ then, $\forall k > 0$ we can write

$$\begin{aligned} P[Y_n > 1/k \text{ i.o.}] &= P[\cap_{m=1}^{\infty} \cup_{n\geq m}^{\infty} \{\omega : Y_n(\omega) > 1/k\}] \\ &\leq P[\cup_{n=k}^{\infty} \{\omega : Y_n(\omega) > 1/k\}] \leq \sum_{n=k}^{\infty} P[\{\omega : Y_n(\omega) > 1/k\}] \\ &\leq \sum_{n=k}^{\infty} 2^{-n} = 2^{-(k+1)}. \end{aligned}$$

Therefore, using the fact that $Y_n \overset{a.s.}{\not\to} 0$ if equivalent to state the existence of a number $k > 0$ such that $P[Y_n > k \text{ i.o}] > 0$, it is possible to write

$$\begin{aligned} P[Y_n \overset{a.s.}{\not\to} 0] &= P[\cup_{k=1}^{\infty} \cap_{m=1}^{\infty} \cup_{n\geq m}^{\infty} \{\omega : Y_n(\omega) > 1/k\}] \\ &= \lim_{k\to\infty} P[\cap_{m=1}^{\infty} \cup_{n\geq m}^{\infty} \{\omega : Y_n(\omega) > 1/k\}] \leq \lim_{k\to\infty} 2^{-(k+1)} = 0. \end{aligned}$$

Hence, $P[\lim_n Y_n \to 0] \to 1$ as we were supposed to show. If the X_n's are not all positive, we can consider $Y_n = |X_n|/a_n$ as, in fact, $Y_n \xrightarrow{a.s.} 0$ iff $|Y_n| \xrightarrow{a.s.} 0$ as $n \to \infty$.

[7] This is possible since $P[X_n > a] = 1 - F_n(a) \to 0$ as $a \to \infty$.

PROBLEM 3.4.7. Let $\{X_n, n \geq 1\}$ be a sequence of independent r.v.'s each of which is normally distributed with mean 0 and variance σ^2. Construct a non decreasing sequence $\{a_n, n \geq 1\}$ such that

$$\limsup_n \frac{X_n}{n} = 1 \ a.s.$$

Then, let $Y_n = \max_{1 \leq k \leq n} X_k$. Is it true that

$$\lim_n \frac{Y_n}{a_n} = 1 \ a.s.?$$

SOLUTION. We begin by stating and proving a lemma.

Lemma. *Let $X \sim N(0, \sigma^2)$. Then, as $x \to \infty$,*

$$P[\{\omega : X(\omega) > x] \approx \frac{\sigma}{x\sqrt{2\pi}} \exp\{-x^2/(2\sigma^2)\}.$$

Proof. The following limit

$$\lim_{x\to\infty} \frac{P[X > x]}{\frac{\sigma}{x\sqrt{2\pi}} \exp\{-x^2/(2\sigma^2)\}}$$

is a $\frac{0}{0}$ form. We can eliminate the indeterminacy using the De L'Hôpital's Rule and Leibnitz's Rule[8]. This gives

$$\begin{aligned}
&\lim_{x\to\infty} \frac{P[X > x]}{\frac{\sigma}{x\sqrt{2\pi}} \exp\{-x^2/(2\sigma^2)\}} \\
&= \lim_{x\to\infty} \frac{\frac{d}{dx}P[X > x]}{\frac{d}{dx}\frac{\sigma}{x\sqrt{2\pi}} \exp\{-x^2/(2\sigma^2)\}} \\
&= \lim_{x\to\infty} \frac{\int_x^\infty \frac{-y}{\sigma^2\sqrt{2\pi\sigma}} \exp\{e^{-y^2/(2\sigma^2)}\}dy - \frac{1}{\sigma\sqrt{2\pi}} \exp\{-x^2/(2\sigma^2)\}}{-\frac{1}{2\sigma\sqrt{2\pi}} \exp\{-x^2/(2\sigma^2)\}(\frac{x^2-\sigma}{x^2})} \\
&= \lim_{x\to\infty} \frac{x^2}{x^2 - \sigma} = 1,
\end{aligned}$$

where we used the fact that $\lim_{x\to\infty} \int_x^\infty \frac{-y}{\sigma^2\sqrt{2\pi\sigma}} \exp\{e^{-y^2/(2\sigma^2)}\}dy = 0.$ □

[8]Leibnitz's Rule for differentiating under the integral sign states that if

$$\phi(x) = \int_{u_1(x)}^{u_2(x)} f(y)dy$$

then,

$$\frac{d}{dx}\phi(x) = \int_{u_1(x)}^{u_2(x)} f'(y)dy + f'(u_2(x))\frac{du_2(x)}{dx} - f'(u_1(x))\frac{du_1(x)}{dx}.$$

Now, let $A_n \equiv \{\omega : X_n(\omega) > n\}$. Since $P[A_n] \to 0$ as $n \to \infty$, we can use the lemma just proved ans state that $\forall \epsilon > 0$, $\exists N_\epsilon$ such that $\forall n > N_\epsilon$

$$(1-\epsilon)\left[\frac{\sigma}{n\sqrt{2\pi}}\exp\{-n^2/(2\sigma^2)\}\right] \le P(A_n) \le (1+\epsilon)\left[\frac{\sigma}{n\sqrt{2\pi}}\exp\{-n^2/(2\sigma^2)\}\right].$$

As the convergence of an infinite sum is not affected by finitely many finite terms, we can easily understand the practical use of the lemma above.

Let $\{a_n, n \ge 1\}$ be an increasing sequence of constants and write

$$B_n = \{\omega : X_n(\omega)/a_n > 1\} = \{\omega : X_n(\omega) > a_n\}.$$

Then, as $n \to \infty$ we have

$$P[B_n] \approx \left[\frac{\sigma}{a_n\sqrt{2\pi}}\exp\{-a_n^2/(2\sigma^2)\}\right]$$

In order to achive $X_n/a_n \overset{a.s.}{\to} 1$, we need to find an increasing sequence of constants $\{a_n, n \ge 1\}$ to satisfy $\sum_{n=1}^{\infty} P[B_n] = \infty$. An educated guess seems to be

$$a_n = \sqrt{(2)\sigma^2(\log n)},\ n = 1, 2, \ldots$$

In fact, with this choice for the a_n's, one can easily verify that when $n \to \infty$, we have

$$P[B_n] \approx \frac{1}{2\sqrt{\pi}n\sqrt{\log n}}.$$

But, $\sum_n P[B_n]$ converges or diverges to ∞ according to whether $\sum_n \frac{1}{\sqrt{2\pi}n\sqrt{\log n}}$ diverges or diverges, respectively. A general result states that the infinite sum $\sum \frac{1}{\sqrt{2\pi}n(\log n)^p}$ converges or diverges according to whether $p > 1$ or $p \le 1$. In our case, $p = 1/2$, the infinite sum diverges and, therefore, the Borel-Cantelli Lemma guarantees that

$$\limsup_n \frac{X_n}{2^{1/2}\sigma(\log n)^{1/2}} \ge 1 \text{ a.s.}$$

If we consider, now, the sequence of events

$$C_n = \{\omega : X_n(\omega)/(2^{1/2}\sigma(\log n)^{1/2}) > 1 + \delta\}$$

for any arbitrary $\delta > 0$, we find that $\sum_{n=1}^{\infty} P[C_n]$ has the same behavior as the infinite sum

$$\sum_{n=1}^{\infty} \frac{1}{(1+\delta)(\log n)^{1/2} n^{(1+\delta)^2}}$$

which is easily seen to converge to a finite limit. Hence, by the Borel Lemma, we can state that $\forall \delta > 0$

$$P[\limsup_n \frac{X_n}{2^{1/2}\sigma(\log n)^{1/2}} \ge 1 + \delta] = 0.$$

Combining the last two results together gives

$$P[\limsup_n \frac{X_n}{2^{1/2}\sigma(\log n)^{1/2}} = 1] = 1.$$

For what attains the second question, we can start by observing that since $Y_n \geq X_n$ $n \geq 1$, it must be $\limsup_n Y_n/a_n \geq 1$ a.s.. We show now that it is actually one. In fact, $\forall \delta > 0$, we have

$$P[\frac{Y_n}{2^{1/2}\sigma(\log n)^{1/2}} \geq (1+\delta)^{1/2}] = \left[1 - \frac{1}{2\sqrt{\pi}(\log n)^{1/2} n^{1+\delta}}\right]^n.$$

Letting $D_n = \{\omega : Y_n(\omega)/(2^{1/2}\sigma(\log n)^{1/2} \leq (1+\delta)^{1/2}\}$, we have that

$$\begin{aligned}\sum_{n=1}^{\infty} P[D_n] &= \sum_{n=1}^{\infty}\left[1 - \frac{1}{2\sqrt{\pi}(\log n)^{1/2} n^{1+\delta}}\right]^n \\ &\geq \sum_{n=1}^{\infty}\left[1 - \frac{1}{n}\right]^n.\end{aligned}$$

The n-th term of this last infinite sum converges to $1/e \neq 0$, hence $\sum_{n=1}^{\infty} P[D_n] = \infty$. The Borel-Cantelli Lemma gives then that for any $\delta > 0$, it must be

$$P[\limsup_n \frac{Y_n}{2^{1/2}\sigma(\log n)^{1/2}} \leq (1+\delta)^{1/2}] = 1.$$

But, we already knew that

$$P[\limsup_n \frac{Y_n}{2^{1/2}\sigma(\log n)^{1/2}} \geq 1] = 1.$$

Combining these last two statements together gives that

$$P[\limsup_n \frac{Y_n}{2^{1/2}\sigma(\log n)^{1/2}} = 1] = 1.$$

In a similar way, we find that for any $n \geq 1$,

$$P[\frac{Y_n}{2^{1/2}\sigma(\log n)^{1/2}} \leq (1-\delta)^{1/2}] = \left(1 - \frac{1}{2\sqrt{\pi}\log(n) n^{1-\delta}}\right)^n.$$

Letting $E_n = \{\omega : Y_n(\omega)/\sqrt{2\sigma^2(\log n)} < (1-\delta)^{1/2}\}$ we can show that $\sum_{n=1}^{\infty} P[E_n] \approx \sum_{n=1}^{\infty} \exp\{-n\delta\} < \infty$. By the Borel-Cantelli Lemma this implies that

$$P[\liminf_n \frac{Y_n}{2^{1/2}\sigma(\log n)^{1/2}} = 1] = 1$$

and the desired conclusion now follows easily.

PROBLEM 3.4.8. Let $\{X_n, n \geq 1\}$ be a sequence of random variables on some probability space $(\Omega, \mathcal{F}, P)$. Prove the following

i. if the X_n's are i.i.d. r.v.'s which are not constant a.e. then,

$$P[\lim_n X_n \textit{ exists}] = 0;$$

ii. if the X_n's, on the contrary, are only assumed to be independent r.v.'s, then

$$P[\lim_n X_n \textit{ exists}] = 0 \textit{ or } 1.$$

SOLUTION. In the first case, if the X_n's are not a.s. constant, it means that there exist two numbers x and y, $x < y$, such that

$$P[X_1 < x] > 0 \; P[X_1 > y] > 0.$$

Then, a simple application of the Borel-Cantelli Lemma gives

$$P[\limsup_n X_n \geq y] = 1 \; P[\liminf_n X_n \leq x] = 1.$$

This means also that

$$P[\liminf_n X_n \leq x < y \geq \limsup_n X_n] = 1$$

and, therefore,

$$P[\lim_n X_n \text{ exists}] = 0$$

i.e., the sequence of random variables $\{X_n, n \geq 1\}$ does not converge with probability 1.

When we drop the assumption that the r.v.'s in the sequence are identically distributed, we need to consider two situations.

1. $\exists x, y$, $x < y$, such that

$$P[\{\omega : X_n(\omega) < x \text{ i.o}\} \cap \{\omega : X_n(\omega) > y \text{ i.o}\}] \geq \delta > 0.$$

If this is the case, the Borel-Cantelli Lemma guarantees that it must be

$$P[\{\omega : X_n(\omega) < x \text{ i.o}\} \cap \{\omega : X_n(\omega) > y \text{ i.o}\}] = 1.$$

This clearly means that

$$P[\{\omega : \liminf_n X_n(\omega) \leq x < y \leq \limsup_n X_n(\omega)\}] = 1$$

and, thus,

$$P[\lim_n X_n \text{ exists}] = 0.$$

2. $\forall x, y$, $x < y$, it is

$$P[\{\omega : X_n(\omega) < x \text{ i.o}\} \cap \{\omega : X_n(\omega) > y \text{ i.o}\}] = 0.$$

In this case, we can write

$$P[\cup_{x,y\in\mathbf{Q}:x<y}\{\omega : X_n(\omega) < x \text{ i.o}\} \cap \{\omega : X_n(\omega) > y \text{ i.o}\}] = 0.$$

Hence, there exists a set of probability 1 on which

$$\liminf_n X_n = \limsup X_n.$$

This means that the sequence $\{X_n, n \geq 1\}$ either converges or diverges to $\pm\infty$. By the Borel-Cantelli Lemma, we find that

$$\begin{aligned} P[X_n \to \infty] &= P[\limsup_n X_n = \infty] = P[\cap_{i=1}^{\infty} X_n > i, \text{ i.o.}] \\ &= \lim_{i\to\infty} P[X_n > i, \text{ i.o.}] = 0 \text{ or } 1. \end{aligned}$$

A similar result holds for $P[X_n \to -\infty]$. Thus, we are in the condition of writing

$$1 = P[\liminf_n X_n = \limsup_n X_n] = P[X_n \text{ converges}] + P[X_n \to \infty] + P[X_n \to -\infty].$$

Finally, since $P[X_n \to \pm\infty] = 0$ or 1, there is only one possible outcome:

$$[X_n \text{ converges}] = 0 \text{ or } 1$$

as we were asked to prove.

3.5 Fatou's Lemma

PROBLEM 3.5.1. Let $(X_1, X_2, \ldots)$ be a sequence of random variables that converge almost surely to a random variable X. Show that if $\sup_n E[X_n^2] < \infty$, then $E[X^2] < \infty$.

SOLUTION. By assumption, $X_n \xrightarrow{a.s.} X$ as $n \to \infty$. Thus, as the function $f(x) = x^2$ is continuous, it follows that $X_n^2 \xrightarrow{a.s.} X^2$ as $n \to \infty$. In addition, $(X_1^2, X_2^2 \ldots)$ is a sequence of $\bar{\mathbf{R}}^+$-valued random variables and hence, by Fatou's Lemma, we can write

$$E[\liminf_n X_n^2] \leq \liminf_n E[X_n^2].$$

As $X_n^2 \xrightarrow{a.s.} X_2$, $\limsup_n X_n^2 = \liminf_n X_n^2 = X^2$ except at most on a set of probability measure 0. On the other hand, we know that

$$\inf_n E[X_n^2] \leq \sup_n E[X_n^2]$$

and, therefore,

$$\liminf_n E[X_n^2] \leq \sup_n E[X_n^2] < \infty.$$

Combining the last three facts together we are led to write

$$E[X^2] = E[\liminf_n X_n^2] \leq \liminf_n E[X_n^2] \leq \sup_n E[X_n^2] < \infty$$

which gives the desired conclusion.

PROBLEM 3.5.2. Let $\{X_n, n \geq 1\}$ be a sequence of r.v.'s defined on some probability space $(\Omega, \mathcal{F}, P)$. Prove that if $X_n \xrightarrow{a.s.} X$ and $E[|X_n|] \to E[|X|] < \infty$ when $n \to \infty$, then

$$E[|X_n - X|] \to 0$$

as $n \to \infty$.

SOLUTION. If we can show that $E[X_n^+]$ and $E[X_n^-]$ converges to $E[X^+]$ and $E[X^-]$, respectively when $n \to \infty$, we can then used Scheffe's Lemma[9] to state that

$$E[|X_n^+ - X^+|] \to 0 \text{ and } E[|X_n^- - X^-|] \to 0$$

as $n \to \infty$. Then, for $n \to \infty$, we can write

$$\begin{aligned} E[|X_n - X|] &= E[|X_n^+ - X_n^- - X^+ + X^-|] \\ &\leq E[|X_n^+ - X^+|] + E[|X_n^- - X^-|] \to 0 \end{aligned}$$

which proves our statement. So, we are left with having to show that it is $\lim_n E[X_n^-] = E[X^-]$ and $\lim_n E[X_n^+] = E[X^+]$. To do this, we start by observing that since $X_n \xrightarrow{a.s.} X$, it is also $X_n^+ \xrightarrow{a.s.} X^+$ and $X_n^- \xrightarrow{a.s.} X^-$ when $n \to \infty$. In addition, we clearly have that $X_n^+, X_n^- \geq 0$ for all $n = 1, 2, \ldots$ This allows us to use Fatou's Lemma for the sequences of r.v.'s $\{X_n^+, n \geq 1\}$ and $\{X_n^-, n \geq 1\}$. This gives that

$$E[X^+] = E[\liminf_n X_n^+] \leq \liminf_n E[X_n^+]$$

and

$$E[X^-] = E[\liminf_n X_n^-] \leq \liminf_n E[X_n^-],$$

respectively. Adding the two inequalities we find also

$$\begin{aligned} E[X^+] + E[X^-] &\leq \liminf_n E[X_n^+] + \liminf_n E[X_n^-] \\ &\leq \liminf(E[X_n^+] + E[X_n^-]) = \lim_n E[|X_n|] \\ &= E[|X|] = E[X^+] + E[X^-]. \end{aligned}$$

[9]Scheffe's Lemma states that for any sequence of r.v.'s $\{X_n, n \geq 1\}$ on a probability space $(\Omega, \mathcal{F}, P)$ such that $X_n \xrightarrow{a.s.} X$, $X_n \geq 0$ $\forall n \geq 1$, and $E[X_n] \to E[X] < \infty$, then

$$E[|X_n - X|] \to 0$$

as $n \to \infty$.

This inequality and the previous one can hold together iff $\liminf_n E[X_n^+] = E[X^+]$ and $\liminf_n E[X_n^-] = E[X^-]$. In addition, it is possible to verify, using the definitions of lim, lim sup and lim inf that

$$\begin{aligned}\limsup_n E[X_n^+] &= \lim_n E[|X_n|] - \liminf_n E[X_n^-] \\ &= \lim_n E[|X_n|] - E[X^-] = E[|X|] - E[X^-] = E[X^+].\end{aligned}$$

A similar reasoning gives $\limsup_n E[X_n^-] = E[X^-]$.
Since we proved that

$$\liminf_n E[X_n^+] = \limsup_n E[X_n^+] = E[X^+]$$

and

$$\liminf_n E[X_n^-] = \limsup_n E[X_n^-] = E[X^-],$$

it follows that $\lim_n E[X_n^+] = E[X^+]$ and $\lim_n E[X_n^-] = E[X^-]$.

PROBLEM 3.5.3. Let $\{X_n, n \geq 1\}$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, P)$ and assume also that there exists a random variable Y on the same probability space such that

$$\sup_n |X_n| \leq Y \textit{ and } E[Y] < \infty.$$

Then,

$$E[\limsup_n X_n] \geq \limsup_n E[X_n].$$

Prove also that this is false if the condition $E[Y] < \infty$ is omitted.

SOLUTION. If we look at the sequence of random variables $\{Y - X_n, n \geq 1\}$, it is clear, because of the assumption $\sup_n |X_n| \leq Y$, that $Y - X_n \geq 0$ for all $n = 1, 2, \ldots$ Thus, we can apply Fatou's Lemma:

$$E[\liminf_n (Y - X_n)] \leq \liminf_n E[Y - X_n].$$

Now, using the definition of limsup, liminf and lim we can write the last inequality as

$$E[Y] - E[\limsup_n X_n] \leq \liminf_n E[Y - X_n].$$

Since, by assumption, $E[Y] < \infty$ we can use the linearity of expectation and rewrite the inequality above as

$$E[Y] - E[\limsup_n X_n] \leq E[Y] - \limsup_n E[X_n].$$

Now, add $-E[Y]$ on both sides to get

$$-E[\limsup_n X_n] \leq -\limsup_n E[X_n].$$

Multiplying both sides by -1 gives the desired conclusion.
To show that the assumption $E[Y] < \infty$ cannot be omitted, consider the following sequence of r.v.'s defined on the probability space $([0,1], \mathcal{B}([0,1]), \lambda)$, λ the Lebesgue measure:

$$X_n(\omega) = \begin{cases} 1 \cdot I_{[0,1]}(\omega) & \text{if } n = 1; \\ (n^2 - n) \cdot I_{(1/n, 1/(n-1)]}(\omega) & \text{if } n \geq 2. \end{cases}$$

It is easy to convince ourselves that there is not random variable Y dominating the sequence $\{X_n, n \geq 1\}$ and such that $E[Y] < \infty$. In fact, such Y should satisfy the following requirements:

$$Y(\omega) \geq (n^2 - n) \cdot I_{(1/n, 1/(n-1)]}(\omega), \ n \geq 2$$

and this implies $E[Y] = \sum_{n=2}^{\infty} 1 = \infty$. Now, $X_n \xrightarrow{a.s.} 0$ as $n \to \infty$ and this implies that $E[\limsup_n X_n] = E[0] = 0$. But, at the same time we have also $E[X_n] = 1$ for any $n = 1, 2, \ldots$ so that $\limsup_n E[X_n] = 1$.

3.6 Weak Convergence

PROBLEM 3.6.1. Let F and F_n, $n = 1, 2, \ldots$, be distribution functions for $\mathbf{R}$. Then $F_n \to F$ as $n \to \infty$ iff there is a dense subset D of $\mathbf{R}$ such that $F_n(x) \to F(x)$ $\forall x \in D$ as $n \to \infty$.

SOLUTION.
($\Rightarrow$) Assume that $F_n \to F$ as $n \to \infty$. By definition, this means that $F_n(x) \to F(x)$ for any $x \in C(F)$, the set of continuity points of F. Since F is a distribution function, it can have at most countably may discontinuity points where $F_n \not\to F$. Hence:

i. $C(F)$ has a countable complement, $C(F)^c$;

ii. every interval is uncountable;

iii. $\Rightarrow \forall x \in C(F)^c$ we can find a sequence $\{y_n\}$, $n = 1, 2, \ldots$, such that $y_n \in C(F)$ and $y_n \to x$ as $n \to \infty$.

Since (iii) tells us that the closure of $C(F)$ is $\mathbf{R}$ we have that $C(F)$ is a dense subset of $\mathbf{R}$. Letting $D = C(F)$ completes the proof of the first implication.

($\Leftarrow$) Assume now that $F_n(x) \to F(x)$ for all $x \in D$, a dense subset of $\mathbf{R}$. Then, $\forall x \in D$ there exists N_x such that $| F_n(x) - F(x) | < \epsilon$ for all $n > N_x$.
Let now $z \in C(F)$; according to the definition of $C(F)$, for any choice of $\epsilon > 0$ there is $\delta > 0$ such that $| F(z) - F(y) | < \epsilon$ for all y for which $| z - y | < \delta$.
Since D is a dense subset of $\mathbf{R}$ we can always find $x_1, x_2 \in D$ such that $z - \delta < x_1 < z < x_2 < z + \delta$.
Now,

1. F and F_n, $n = 1, 2, \ldots$, are distribution functions and hence they are non-decreasing;

2. $\mid F_n(x_1) - F_n(x_1) \mid < \epsilon \; \forall n > N_1$; similarly $\mid F_n(x_2) - F(x_2) \mid < \epsilon \; \forall n > N_2$;

3. letting $N = \max\{N_1, N_2\}$ we have that $\forall n > N$,

$$F(x_1) - \epsilon < F_n(x_1) \leq F_n(z) \leq F_n(x_2) < F(x_2) + \epsilon;$$

4. $z \in C(F)$ and $\mid x_1 - z \mid < \delta$, $\mid x_2 - z \mid < \delta$ and hence, according to what we stated above this means that $F(z) - 2\epsilon < F(x_1) - \epsilon$ and $F(x_2) + \epsilon < F(z) + 2\epsilon$. Therefore we can write

$$F(z) - 2\epsilon < F(x_1) - \epsilon < F_n(x_1) \leq F_n(z) \leq F_n(x_2) < F(x_2) + \epsilon < F(z) + 2\epsilon;$$

5. since ϵ is arbitrary, it follows that $F_n(z) \to F(z) \; \forall z \in C(F)$.

This last result implies that $F_n \to F$ as $n \to \infty$ and, hence, the proof of the second implication.

PROBLEM 3.6.2. Let $X_1, X_2, \ldots$, be i.i.d. $\mathbf{R}$-valued random variables with distribution function F, and let F_n denote the empirical distribution function of $X_1, X_2, \ldots$, That is, for any Borel set A

$$F_n(A) = \frac{1}{n} \sum_{i=1}^{n} I_A(X_i).$$

Show that F_n converges weakly to F almost surely as $n \to \infty$.

SOLUTION. It is easily seen that for any choice of x we have

$$F_n(x) \to F(x) \;\; a.s.$$

as $n \to \infty$. This is a simple consequence of the SLLN. In fact, for any Borel set A it is easily checked that $I_A(X_i)$, $i = 1, 2, \ldots$, form a sequence of i.i.d. $\mathbf{R}$-valued r.v's each with mean $F(A)$ and variance $F(A)[1 - F(A)]$. Since $F(A)$ is finite the SLLN applies to give $F_n(A) \to F(A)$ almost sureley as $n \to \infty$. If we let $A = (-\infty, x]$ then it follows that $F_n(x) \to F(x)$ almost surely as $n \to \infty$. Since x is an arbitrary point of $\mathbf{R}$ this is true for all rational points, that is:

$$P[\{\omega : F_n(q, \omega) \to F(q, \omega), \;\; q \in \mathbf{Q}\}] = 1.$$

This implies that $F_n \to F$ almost surely as $n \to \infty$ on a dense subset of the real line and, using the previous problem this implies that $F_n \to F$ weakly almost surely as $n \to \infty$.

PROBLEM 3.6.3. Let X and X_n, $n = 1, 2, \ldots$, be $\mathbf{Z}$−valued random variables with probability mass functions $p(k)$ and $p_n(k)$. Show that X_n converges to X in distribution iff $p_n(\{k\})$ converges to $p(k)$ for all integers k.

SOLUTION.
($\Rightarrow$) Assume that $X_n \xrightarrow{D} X$ and let Q_n and Q be probability measures on $\mathbf{R}$ which correspond to X_n and X, respectively. Then, according to the definition of convergence in distribution, this means that $Q_n \to Q$ when $n \to \infty$. Since we are dealing with $\mathbf{Z}$−valued random variables, a special case of $\mathbf{R}$−valued random variables, we can apply the Portmanteau Theorem. This implies that $Q_n(A) \to Q(A)$ as $n \to \infty$ for any Borel subset A of the real line for which $Q(\partial A) = 0$.
If we consider the following Borel subsets of $\mathbf{R}: A_k = (k - 1/2, k + 1/2)$ $k \in \mathbf{Z}$, it is easily seen that $Q(\partial A_k) = 0$ and $Q(A_k) = \sum_{k \in \mathbf{Z} \cap A_k} p(\{k\}) = p(\{k\})$. In the same way we can check that $Q_n(A_k) = p_n(\{k\})$. We know, by assumption, that $X_n \xrightarrow{D} X$ and hence $Q_n(A) \to Q(A)$ for all Borel subset A whose boundary set has measure zero. This implies $Q_n(A_k) \to Q(A_k)$ which in turn implies $p_n(\{k\}) \to p(\{k\})$ for any integer k.

($\Leftarrow$) Assume now that $p_n(k) \to p(k)$ for any $k \in \mathbf{Z}$ as $n \to \infty$. Let $H = \{k \in \mathbf{Z} : p(k) > p_n(k)\}$ and let $M = [-m, m] \cap \mathbf{Z}$. Now, for any $\epsilon > 0$, we can always find m such that $\sum_{k \in M^c} p(k) < \epsilon$.
If ν is the counting measure on $\mathbf{Z}$, we can write $\sum_k f(k)p_n(k) = \int_{\mathbf{Z}} f(k)p_n(k)d\nu$ and, equivalently, $\sum_k f(k)p(k) = \int_{\mathbf{Z}} f(k)p(k)d\nu$. If we can prove that

$$\sum_k f(k)p_n(k) \to \sum_k f(k)p(k)$$

for any bounded and continuous function f we are done since we know that this is equivalent to showing that we have convergence in distribution.
It is easily checked that we can write

$$\left| \sum_k f(k)p_n(k) - \sum_k f(k)p(k) \right| \leq \sum_k | f(k) || p_n(k) - p(k) | \leq$$

$$\sup_{k \in \mathbf{Z}} | f(k) | \sum_k | p_n(k) - p(k) | \leq$$

$$\sup_{k \in \mathbf{Z}} | f(k) | [\sum_{k \in H^c} (p_n(k) - p(k)) + \sum_{k \in H} (p(k) - p_n(k)] =$$

$$= 2 \sup_{k \in \mathbf{Z}} | f(k) | \sum_{k \in H} (p(k) - p_n(k)) =$$

$$2 \sup_{k \in \mathbf{Z}} | f(k) | \sum_{k \in H \cap M} (p(k) - p_n(k)) + 2 \sup_{k \in \mathbf{Z}} | f(k) | \sum_{k \in H \cap M^c} (p(k) - p_n(k)).$$

Now, the function f is continuous and bounded, therefore its supremum is finite. Besides, the first summation involves only a finite number of terms so we can interchange limit and summation and since $p_n(k) \to p(k)$ for all $k \in H \cap M$ the first summation goes to zero when $n \to \infty$. The last summation goes to zero as well; in fact:

$$\sum_{k \in H \cap M^c} p(k) - p_n(k) \leq \sum_{k \in M^c} p(k) < \epsilon$$

and since ϵ is arbitrary, the conclusion follows.

PROBLEM 3.6.4. Suppose F is a continuous distribution function and F_n is a sequence of distribution functions that converges pointwise to F, that is

$$F_n(x) \to F(x)$$

for all x. Show that F_n converges uniformly to F and suggest the relevance of this fact in the applications of the CLT.

SOLUTION. Since F is a distribution function, for any choice of $\epsilon > 0$, we can always find two real numbers, a and b with $a < b$ such that $F(a) < \epsilon/2$ and $F(b) > 1 - \epsilon/2$. As, by assumption, F is continuous and $[a, b]$ is a compact subset of the real line, it follows that F is uniformly continuous on $[a, b]$. We can therefore choose $\delta > 0$ such that $\mid F(x) - F(y) \mid< \epsilon/2$ if $x, y \in [a, b]$ and $\mid x - y \mid< \delta$. Consider now the following finite sequence

$$x_1 = a \leq x_2 \leq x_3 \leq \ldots \leq x_k = b$$

with $x_{k+1} - x_k < \delta$ for $k = 1, 2, \ldots, K - 1$. Since $F_n(x_k) \to F(x_k)$ for all k, we can choose an integer N such that for $n \geq N$ we have $\mid F_n(x_k) - F(x_k) \mid \epsilon/2$ for all $k = 1, 2, \ldots, K - 1$. This is possible because we are considering a finite sequence and so we can choose N. If $n \geq N$, and $x \leq a$, monotonicityof F and F_n imply that

$$0 \leq F_n(x) \leq F_n(a)$$

and

$$0 \leq F(x) \leq F(a).$$

The definitions of N and a imply that

$$F_n(a) < F(a) + \epsilon/2 < \epsilon$$

and

$$F(a) < \epsilon/2 < \epsilon.$$

These last four facts considered together imply the following

$$\mid F_n(x) - F(x) \mid< \epsilon$$

for any $x \leq a$. Similarly, for $x \geq b$ we find

$$1 \geq F_n(x) \geq F_n(b) > F(b) - \epsilon/2 > 1 - \epsilon$$

$$1 \geq F(x) > 1 - \epsilon,$$

and, again, $| F_n(x) - F(x) | < \epsilon$.
Finally, suppose $x_k \leq x \leq x_{k+1}$ for some $k = 1, 2, \ldots, K-1$. In this case we have

$$F(x) - \epsilon < F(x_k) - \epsilon/2 < F_n(x_k) \leq F_n(x)$$

and

$$F_n(x) \leq F_n(x_{k+1}) < F(x_{k+1}) + \epsilon/2 < F(x) + \epsilon.$$

So $| F_n(x) - F(x) | < \epsilon$ in this case as well.
Since any $x \in \mathbf{R}$ falls in one of these three cases, this shows that

$$| F_n(x) - F(x) | < \epsilon$$

for all x and for any $n > N$. We have therefore proved that F_n converges uniformly to F.
The relevance for the CLT is that in the CLT the sequence of distribution functions F_n converges to the distribution of a normal random variable which is a continuous one. Therefore the result just proved above tells us that the convergence of the distribution functions involved is of uniform type.

PROBLEM 3.6.5. [10] Suppose X, Y, X_n, and Y_n, $n = 1, 2, \ldots$, are $\mathbf{R}-$ valued random variables and assume that X_n converges in distribution to X and Y_n converges in distribution to Y.

(a) Suppose that Y is almost surely equal to a constant c. Show that (X_n, Y_n) converges to (X, c).

(b) Suppose that X and Y are independent and for each n the random variables X_n and Y_n are also independent. Show that (X_n, Y_n) converges in distribution to (X, Y).

(c) Give a counter example to show that (X_n, Y_n) need not converge in distribution to (X, Y) if the assumption in (a) or (b) do not hold.

SOLUTION.
(a). As, by assumption $Y = c$ a.s., this means that $Y_n \xrightarrow{P} c$ a.s. as $n \to \infty$. In our case we have also that $X_n \xrightarrow{D} X$ and therefore we have that for any real constant, a,

$$aX_n \xrightarrow{D} aX \quad \text{and} \quad aY_n \xrightarrow{P} ac, \ a.s$$

when the multiplication of a random variable by a constant is looked at as a continuous function in the random variable itself. Slutsky's Theorem is now enough to state that for arbitrary real constants, a and b, we have

$$aX_n + bY_n \xrightarrow{D} aX + bc.$$

[10]This problem is taken from Billingsley, Measure and Probability, 1985.

Given this fact, the Cramér and Wold device[11] is all we need to state that

$$(X_n, Y_n) \xrightarrow{D} (X, c)$$

as $n \to \infty$.

(b) To prove the second statement a couple of preparatory lemmas are called for.

Definition: For random vectors $X_1, X_2, \ldots,$ and X we say that $X_n \xrightarrow{P} X$ if $\| X_n - X \| \xrightarrow{P} 0$, where $\| z \| = (\sum_{i=1}^T z_i^2)^{1/2}$.

Lemma 1. *For random vectors in* $\mathbf{R}^T$, $X_1, X_2, \ldots,$ *and* X, $X_n \xrightarrow{P} X$ *iff the corrispondent component-wise convergences hold.*

Proof. Assume that $X_n \xrightarrow{P} X$, then according to the definition there exist arbitrary positive numbers ϵ and ν such that

$$P[\| X_n - X \| > \epsilon] < \nu \quad \Rightarrow \quad P[(\sum_{i=1}^T (X_{nk} - X_k)^2)^{1/2} > \epsilon] < \nu \quad \forall n > N.$$

This then implies that for any $k = 1, 2, \ldots, T$ and for $n > N$

$$P[\sum_{i=1}^T (X_{nk} - X_k)^2 > \epsilon^2] < \nu \quad \forall n > N \quad \Rightarrow P[| X_{nk} - X_k | > \epsilon] < \nu$$

and hence $X_{nk} \xrightarrow{P} X_k$ as $n \to \infty$ for any $k = 1, 2, \ldots, T$.

Assume now that $X_{nk} \xrightarrow{P} X_k$ for all $k = 1, 2, \ldots, T$. By definition,this means that $P[| X_{nk} - X_k | > \epsilon] < \nu$ for all $n > N$ and $k = 1, 2, \ldots, T$. As addition is a continuous function we have also $\sum_{i=1}^T X_{nk} \xrightarrow{P} \sum_{i=1}^T X_k$ and therefore there exist ϵ and ν such that

$$\begin{aligned}
\nu > P[| \sum_{i=1}^T X_{nk} - \sum_{i=1}^T X_k | > \epsilon] &= P[| \sum_{i=1}^T X_{nk} - X_k |] = P[\sum_{i=1}^T (X_{nk} - X_k)^2 > \epsilon^2] \\
&= P[\sum_{i=1}^T (X_{nk} - X_k)^2 + 2 \sum_h \sum_k | X_{nh} - X_h || X_{nk} - X_k | > \epsilon^2] \\
&\geq P[\sum_{i=1}^T (X_{nk} - X_k)^2 > \epsilon^2] = P[(\sum_{i=1}^T (X_{nk} - X_k)^2)^{1/2} > \epsilon] = P[\| X_n - X \| > \epsilon]
\end{aligned}$$

and therefore, as ϵ and ν are arbitrary, $X_n \xrightarrow{P} X$ as $n \to \infty$. □

[11] The **Cramér and Wold device** (see e.g. Billingsley, 1985, Thm 29.4 for a proof) states that for random vectors $X_n = (X_{n1}, X_{n2}, \ldots, X_{nk})$ and $X = (X_1, X_2, \ldots, X_k$ a necessary and sufficient condition for X_n to converge in distribution to X is that

$$\sum_{j=1}^k \lambda_j X_{nj} \xrightarrow{D} \sum_{j=1}^k \lambda_j X_j$$

for each vector $(\lambda_1, \lambda_2, \ldots, \lambda_k)$ in $\mathbf{R}^k$.

Lemma 2. *Let P_n and Q_n, P and Q be probability measures on* $\mathbf{R}$. *Then* $P_n \times Q_n \to P \times Q$ *iff* $P_n \to P$ *and* $Q_n \to Q$ *when* $n \to \infty$.

Proof. If $P_n \times Q_n \to P \times Q$ when $n \to \infty$, we know that this implies (because of the Portmanteau Theorem) that, for example

$$\limsup_{n\to\infty} P_n \times Q_n(C) \leq P \times Q(C)$$

for all closed subset of $\mathbf{R}^2$. So, let $C = A \times \mathbf{R}$ where A is a closed subset of the real line. Then C is a closed subset in $\mathbf{R}^2$ and hence

$$P_n \times \limsup_{n\to\infty} Q_n(A \times \mathbf{R}) = \limsup_{n\to\infty} P_n(A) \cdot Q_n(\mathbf{R}) =$$
$$= \limsup_{n\to\infty} P_n(A) \leq P(A) \times Q(\mathbf{R}) = P(A).$$

and since A is an arbitrary closed subset of the real line it follows by the same Portmanteau Theorem that $P_n \to P$ as $n \to \infty$. The proof that $Q_n \to Q$ as $n \to \infty$ is exactly the same when sets like $\mathbf{R} \times A$ are considered.
To prove the reverse implication let's start assuming that $P_n \to P$ and $Q_n \to Q$ as $n \to \infty$. Then, Skorokhod's Theorem guarantees that there exist probability spaces $(\Omega_X, \mathcal{F}_X, P_X)$ and $(\Omega_Y, \mathcal{F}_Y, P_Y)$ and random variables X_n and X, Y_n and Y definedon Ω_X and Ω_Y, respectively, such that

$$X_n \sim P_n,\ \ Y_n \sim Q_n,\ \ , X \sim P \text{ and } Y \sim Q$$

such that $X_n \overset{a.s.}{\to} X$ and $Y_n \overset{a.s.}{\to} Y$ as $n \to \infty$. Hence, by Lemma 1, we know that this implies $(X_n, Y_n) \overset{P}{\to} (X, Y)$ and since convergence in probability implies convergence in distribution, we have also that $(X_n, Y_n) \overset{D}{\to} (X, Y)$ as $n \to \infty$. Since $P_n \times Q_n$ is the probability distribution of (X_n, Y_n) and $P \times Q$ is the probability distribution of (X, Y), this fact implies that $P_n \times Q_n \to P \times Q$ as $n \to \infty$. □

In our case, we know by assumption, that $X_n \overset{D}{\to} X$ and $Y_n \overset{D}{\to} Y$ and we also know that the X_n's and the Y_n's are independent as are X and Y.These two facts together imply that $P_n \to P$ and $Q_n \to Q$ when $n \to \infty$, P_n and Q_n being the distribution of X_n and Y_n, respectively. The assumptions of independence implies that $P_n \times Q_n$ is the distribution of (X_n, Y_n) while $P \times Q$ is that of (X, Y). Lemma 2 is now enough to conclude that $P_n \times Q_n \to P \times Q$ as $n \to \infty$ and therefore

$$(X_n, Y_n) \overset{D}{\to} (X, Y)$$

as $n \to \infty$.

(c) To show that the conditions in (a) and (b) cannot be avoided we can consider the following counter example. Let $X \sim N(0,1)$, $Y \sim N(0,1)$ and assume also that X and Y are independent. Then, let $X_n = Y_n \sim N(0,1)$. Clearly X_n and Y_n are not independent and $X_n \overset{D}{\to} X$, $Y_n \overset{D}{\to} Y$, but $(X_n, Y_n) \not\to (X, Y)$ as X and Y are independent.

PROBLEM 3.6.6. Let F_n, $n = 1, 2, \ldots$, F and G be distribution functions for $\mathbf{R}$. Suppose that $F_n \to F$ and $F_n \to G$ as $n \to \infty$. Then $F = G$.

SOLUTION. Clearly, $F(x) = G(x)$ at any point $x \in C(F) \cap C(G)$. This is the consequence of assuming that $F_n \to F$ and $F_n \to G$ together with the definition of convergence in distribution. The set of points where F and G are not continuous is at most countable, therefore $C(F) \cap C(G)$ is a dense subset of the real line as it has a countable complement. This tells us that for any $y \in \mathbf{R}$ we can find a sequence of points $\{x_n, n = 1, 2, \ldots\}$ in $C(F) \cap C(G)$ such that $\lim_n x_n \searrow y$ as $n \to \infty$. Assume now that y is a generic point in $(C(F) \cap C(G))^C$, we want to prove that $F(y) = G(y)$.
As $C(F) \cap C(G)$ is a dense subset of the real line, we can find a sequence of points $\{x_n, n = 1, 2, \ldots,$ in $C(F) \cap C(G)$ such that $x_n \searrow y$ as $n \to \infty$. We clearly have:

$$| F(y) - F(x_n) | < \epsilon/4 \quad \forall n > N_1$$

as $x_n \searrow y$ and F being a distribution function is right continuous;

$$| F(x_n) - F_m(x_n) | < \epsilon/4 \quad \forall m > M_1$$

as $F_n \to F$ and $x_n \in C(F)$;

$$| F_m(x_n) - G(x_n) | < \epsilon/4 \quad \forall m > M_2$$

as $F_n \to G$ and $x_m \in C(G)$ and

$$| G(x_n) - G(y) | < \epsilon/4 \quad \forall n > N_2$$

as $x_n \searrow y$ and G being a distribution function is right continuous. These four facts together tell us that for all $n, m > \max\{N_1, N_2, M_1, M_2\}$ we have

$$| F(y) - G(y) | \leq | F(y) - F(x_m) | + | F(x_m) - F_n(x_m) | +$$

$$+ | F_n(x_m) - G(x_m) | + | G(x_m) - G(y) | < \epsilon$$

and since y is arbitrary we are done.

PROBLEM 3.6.7. In either the $\mathbf{R}$ or $\bar{\mathbf{R}}$ setting, suppose that $\{Q_n : n = 1, 2, \ldots\}$ is a sequence of probability distributions that has the property that, for some probability distribution Q, every sequence has a further subsequence that converges to Q. Then $Q_n \to Q$ as $n \to \infty$.

SOLUTION. To prove this result we need two basic lemmas from General Topology.

Lemma 1. *Every bounded sequence $\{x_n\}$ of real numbers contains a convergent subsequence.*

Proof. The proof is just an application of the Bolzano Weierstrass Theorem. □

Lemma 2. *If $\{x_n\}$ is a sequence of real numbers that, as a set, is relatively sequentially compact and every convergent subsequence has the same limit,then $\{x_n\}$ also converges to that limit.*

Proof. The proof follows easily from the fact that $\mathbf{R}$ or $\bar{\mathbf{R}}$ with the usual metrics are Hausdorff spaces. □

For any $x \in \mathbf{R}$ or $\bar{\mathbf{R}}$ we observe that

$$\{F_n(x) : n = 1, 2, \ldots\}$$

is a bounded sequence of real numbers thus, by Lemma 1, has a convergent subsequence and, therefore, it is a relatively sequentially compact set.
Let now $\{Q_n : n = 1, 2, \ldots, \}$ be a sequence of probability measures which, by assumption, has a subsequence $\{Q_{nk} : k = 1, 2, \ldots, \}$ which has a further subsequence $\{Q_{nkh} : h = 1, 2, \ldots, \}$ such that $Q_{nkh} \to Q$ as $h \to \infty$.
Using the definition of convergence in distribution, this means that $F_{nkh}(x) \to F(x)$ $\forall x \in C(F)$ where F_{nkh} and F are the distribution functions associated with Q_{nkh} and Q, respectively.
Since we know that $\{F_n(x) : n = 1, 2, \ldots\}$ is a relatively sequentially compact set for any choice of $x \in C(F)$, Lemma 2 implies that $F_{nk}(x) \to F(x)$ as $k \to \infty$.
We can now use the same argument to show that $F_n(x) \to F(x)$ as $n \to \infty$ for any $x \in C(F)$ as well.
By definition of convergence in distribution, this implies that $Q_n \to Q$ as $n \to \infty$ and our proof is therefore complete.

3.7 Laws of Large Numbers

PROBLEM 3.7.1. Let $\{X_n : n \geq 1\}$ be a sequence of i.i.d. random variables, uniformly distributed on the interval $[0, L]$ and let $f(\cdot)$ be a real periodic function with period L, continuous on $\mathbf{R}$. Prove that for every $u \in \mathbf{R}$

$$\frac{1}{n} \sum_{i=1}^{n} f(u + X_i) \xrightarrow{P} \int_0^L f(u)\, du \quad \text{as } n \to \infty.$$

SOLUTION. It is useful to start by recalling a few basic facts about periodic functions. This is done in the following

Lemma. *Let $f(\cdot)$ be periodic, of period L. Suppose that for some w_0 the integral $\int_{w_0}^{w_0+L} f(v)\, dv$ exists. Then, for every value of r the integral $\int_{w_0}^{w_0+L} f(v + r)\, dv$ exists. For every w_1 the integral $\int_{w_1}^{w_1+L} f(v)\, dv$ exists as well. In addition, all the integrals have the same value.*

Proof. See e.g. R.K. Ritt, Fourier Series, 1970; Theorem 1, p. 13. □

Now, let $Y_n \equiv u + X_n$, $n = 1, 2, \ldots$ Then, it is easily seen that $\{Y_n : n \geq 1\}$ is a sequence of i.i.d. $\sim U(u, u+L)$ distributed random variables. Using the Lemma, we found also that

$$E[Y_1] = \int_u^{u+L} f(u+y)\,dy = \int_u^{u+L} f(y)\,dy = \int_0^L f(y)\,dy.$$

Since, by assumption, $f(\cdot)$ is continuous and $[0, L]$ is a compact interval, it is easily seen that $f(\cdot)$ is bounded on $[0, L]$. This implies that $E[|Y_1|] < \infty$. So, using Khintchine's Law of Large Numbers, we have the desired conclusion, i.e.:

$$\frac{1}{n}\sum_{i=1}^{n} f(u + X_k) \xrightarrow{P} \int_0^L f(u)\,du$$

when $n \to \infty$.

PROBLEM 33. Let f, g be two real-valued functions such that $0 \leq f(x) \leq cg(x)$ for every $x \in [0,1]$ and for some $c > 0$. Compute

$$\lim_n \underbrace{\int_0^1 \int_0^1 \cdots \int_0^1}_{n-times} \frac{f(x_1) + f(x_2) + \ldots + f(x_n)}{g(x_1) + g(x_2) + \ldots + g(x_n)}\,dx_1\,dx_2 \ldots dx_n.$$

SOLUTION. If $\{X_n : n \geq 1\}$ is a sequence of i.i.d. $U(0,1)$ random variables, and we let

$$H_n(x_1, x_2, \ldots, x_n) = \frac{f(x_1) + f(x_2) + \ldots + f(x_n)}{g(x_1) + g(x_2) + \ldots + g(x_n)}$$

it is easily seen that the quantity we want to compute the limit for when $n \to \infty$ is the same as $E[H_n(X_1, X_2, \ldots, X_n)]$. Now, because of the assumptions about $\{X_n : n \geq 1\}$, $f(\cdot)$, and $g(\cdot)$, we have in first place that $E[f(X_1)] = \int_0^1 f(y)\,dy$, and $E[g(X_1)] = \int_0^1 g(y)\,dy$; and we also have that $E[|f(X_1)|] < \infty, E[|g(X_1)|] < \infty$. Then, using Khintchine's Law of Large Numbers, it follows that

$$H_n(X_1, X_2, \ldots, X_n) = \frac{\frac{1}{n}\sum_{i=1}^n f(X_i)}{\frac{1}{n}\sum_{i=1}^n g(X_i)} \xrightarrow{P} \frac{\int_0^1 f(y)\,dy}{\int_0^1 g(y)\,dy} \equiv H$$

as $n \to \infty$. Using the assumption that $0 \leq f(x) \leq cg(x)$ for every $x \in [0,1]$ and for some $c > 0$, we have also that $H_n(X_1, X_2, \ldots, X_n) \leq c$ and hence, using the Dominated Convergence Theorem (see the remark below), we have the following

$$\lim_n E[H_n] = E[H] = \frac{\int_0^1 f(y)\,dy}{\int_0^1 g(y)\,dy}$$

which is also the limit we were trying to compute.
Remark. The usual statement of Lebesgue DCT requires that the sequence of r.v.'s $\{X_n : n \geq 1\}$ is such that $X_n \xrightarrow{P} X$ when $n \to \infty$ in order for us to conclude that $E[X_n] \to E[X]$ as $n \to \infty$. We just concluded above that $X_n \xrightarrow{P} X$ only. Nevertheless, we have the following:

Lemma. *Let $\{X_n : n \geq 1\}$ be a sequence of real-valued random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ and let X be another real-valued random variable defined on the same probability space and such that $X_n \xrightarrow{P} X$ as $n \to \infty$. If $|X_n| \leq Y$ and $E[Y] < \infty$, then $E[X_n] \to E[X]$ as $n \to \infty$.*

Proof. If $X_n \xrightarrow{P} X$ as $n \to \infty$, then every subsequence $\{X_{n_m} : n_m \geq 1\}$ has a further subsequence $\{X_{n_{m_k}} : n_{m_k} \geq 1\}$ such that $X_{n_{m_k}} \xrightarrow{a.s.} X$. Using Lebesgue's DCT we have that $E[X_{n_{m_k}}] \to E[X]$. Finally, one uses the fact that if $\{y_n : n \geq 1\}$ is a sequence of elements of a topological space and every subsequence $\{y_{n_m} : n_m \geq 1\}$ has a further subsequence $\{y_{n_{m_k}} : n_{m_k} \geq 1\}$ that converges to y then $y_n \to y$ as $n \to \infty$. Applying this result to the sequence $\{E[X_n] : n \geq 1\}$ completes the proof. □

PROBLEM 3.7.2. [12]. Compute the following limit

$$\lim_n \underbrace{\int_0^1 \int_0^1 \cdots \int_0^1}_{0 \leq x_i \leq 1, 1 \leq i \leq n, x_1^2 + x_2^2 + \ldots + x_n^2 \leq \sqrt{n}} dx_1\, dx_2 \ldots dx_n$$

SOLUTION. If $\{X_n : n \geq 1\}$ is a sequence of i.i.d. $U(0,1)$ random variables, the problem is easily seen to be equivalent to computing

$$\lim_n \Pr\left[\sum_{i=1}^n X_i^2 \leq \sqrt{n}\right] = \lim_n \Pr\left[\frac{1}{n}\sum_{i=1}^n X_i^2 \leq \frac{1}{\sqrt{n}}\right].$$

Now, $\{X_n^2 : n \geq 1\}$ is a sequence of i.i.d. random variables such that $E[X_1^2] = 1/3$ and, therefore, a simple application of Khintchine's Law of Large Numbers gives

$$\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{P} \frac{1}{3} \quad \text{as } n \to \infty.$$

Then, it is

$$\lim_n \underbrace{\int_0^1 \int_0^1 \cdots \int_0^1}_{0 \leq x_i \leq 1, 1 \leq i \leq n, x_1^2 + x_2^2 + \ldots + x_n^2 \leq \sqrt{n}} dx_1\, dx_2 \ldots dx_n = \Pr[1/3 \leq 0] = 0.$$

[12]This problem is taken from A.Ya Dorogovtsev, D.S. Silvestrov, A.V. Skorohod and M.I, Yadrenko, Probability Theory: Collection of Problems, American Mathematical Society, 1991; Problem III.4.10.

PROBLEM 3.7.3. Let $\{X_n : n \geq 1\}$ be a sequence of i.i.d. $N(0,1)$ random variables. Let $Y_i = \sin(X_i/X_{i+1})$, i = 1, 2, ... Let

$$Z_n = \frac{Y_1 + Y_3 + \ldots + Y_{2n-1}}{n} \text{ and } T_n = \frac{Y_1 + Y_2 + \ldots + Y_n}{n}.$$

Does Z_n converges in probability as $n \to \infty$? Does T_n?

SOLUTION. It is easy to show that $W_i \equiv X_i/X_{i+1} \sim \text{Cauchy}(0,1)$, $i = 1, 2, \ldots$ and that $Y_i = \sin(W_i)$, $i = 1, 2, \ldots$ are identically distributed r.v.'s having zero mean. In fact,

$$E[Y_1] = \int_{-\infty}^{\infty} \frac{\sin(y)}{\pi(1+y^2)}\, dy$$

and the integrand function is odd. In addition, because of the assumption that the X_i's are i.i.d. random variables, the sequence $\{Y_{2n-1} : n \geq 1\}$ is also an i.i.d. sequence of random variables for which $E[|Y_1|] < \infty$. Then it is possible to use Khintchine's Law of Large Numbers and state that

$$T_n \xrightarrow{P} 0 \text{ as } n \to \infty.$$

The sequence $\{Y_n : n \geq 1\}$ is not i.i.d. so we cannot use Khintchine's Law of Large Numbers to prove convergence in probability for T_n. However, we notice that for the sequence $\{Y_{2n} : n \geq 1\}$ a result similar to the one proved above holds, i.e.:

$$\frac{Y_2 + Y_4 + \ldots + Y_{2n}}{n} \xrightarrow{P} 0 \text{ as } n \to \infty.$$

If we let $V_n = \frac{Y_2 + Y_4 + \ldots + Y_{2n}}{n}$ it is possible to write

$$\begin{aligned} T_n &= \frac{Y_1 + Y_3 + \ldots + Y_{2\cdot\lfloor\frac{n+1}{2}\rfloor - 1}}{n} + \frac{Y_2 + Y_4 + \ldots Y_{2\cdot\lfloor\frac{n}{2}\rfloor}}{n} \\ &= \frac{Y_1 + Y_3 + \ldots + Y_{2\cdot\lfloor\frac{n+1}{2}\rfloor - 1}}{\lfloor\frac{n+1}{2}\rfloor} \cdot \frac{\lfloor\frac{n+1}{2}\rfloor}{n} + \frac{Y_2 + Y_4 + \ldots Y_{2\cdot\lfloor\frac{n}{2}\rfloor}}{\lfloor\frac{n}{2}\rfloor} \cdot \frac{\lfloor\frac{n}{2}\rfloor}{n} \end{aligned}$$

where $\lfloor\cdot\rfloor$ is the function that once applies to x returns the integer value closest to x. Thus,

$$T_n = Z_{m(n)} \cdot \frac{\lfloor\frac{n+1}{2}\rfloor}{n} + V_{p(n)} \cdot \frac{\lfloor\frac{n}{2}\rfloor}{n}$$

and $m(n), p(n)$ are such that $m(n), p(n) \uparrow \infty$ as $n \to \infty$. Since we know that $V_{p(n)}, Z_{m(n)} \xrightarrow{P} 0$ as $m(n), p(n) \to \infty$ and $\frac{\lfloor\frac{n+1}{2}\rfloor}{n}, \frac{\lfloor\frac{n}{2}\rfloor}{n} \to \frac{1}{2}$ as $n \to \infty$ it follows that $T_n \xrightarrow{P} 0$ as $n \to \infty$.

PROBLEM 3.7.4. Let $\{X_n : n \geq 1\}$ be a sequence of i.i.d $U(0,1)$ random variables. Define

$$Y_n = \left(e^n \cdot \Pi_{i=1}^n X_i\right)^{1/n}.$$

Compute the limit in probability for Y_n as $n \to \infty$.

SOLUTION. Let $S_n = \log(Y_n)$ for $n = 1, 2, \ldots$; i.e.:

$$S_n = \log(Y_n) = 1 + \frac{1}{n}\sum_{i=1}^{n} \log(X_i).$$

Now, $E[\log(X_1)] = -1$ and $E[|\log(X_1)|] < \infty$ so that one can use Khintchine's Law of Large Numbers to conclude that $S_n \xrightarrow{P} 0$ as $n \to \infty$. Then, $Y_n \xrightarrow{P} 1$ as $n \to \infty$.

PROBLEM 3.7.5. Let $\{X_n : n = 1, 2, \ldots\}$ be a sequence of independent random variables defined on a probability space $(\Omega, \mathcal{F}, P)$. Let also Y be a random variable with finite variance and such that $E[Y] = \alpha$. Prove that

$$\frac{1}{n}\sum_{i=1}^{n} E[Y \mid X_i] \xrightarrow{a.s.} \alpha.$$

SOLUTION. First we prove the following

Lemma. *Let X and Y be arbitrary $\mathbf{R}$-valued random variables and $\mathcal{H}$, $\mathcal{G}$ two independent σ-algebras. Then $E[X \mid \mathcal{H}]$ and $E[Y \mid \mathcal{G}]$ are independent.*

Proof. Let B_1 and B_2 be arbitrary Borel sets. Using the definition of conditional expectation, the events $\{E[X \mid \mathcal{H}] \in B_1\}$ and $\{E[Y \mid \mathcal{G}] \in B_2\}$ belong to $\mathcal{H}$ and $\mathcal{G}$, respectively. Since $\mathcal{H}$ and $\mathcal{G}$ are independent, so are $E[X \mid \mathcal{H}]$ and $E[Y \mid \mathcal{G}]$. □

Using the Lemma just proved, it is easily checked that $\{E[X \mid X_i] : i = 1, 2, \ldots\}$ represents a sequence of independent random variables. In addition, we have

$$\mathrm{Var}[E[Y \mid X_i]] \le \mathrm{Var}[Y] < \infty$$

and

$$E[E[Y \mid X_i]] = E[Y] = \alpha.$$

Then the statement of the problem follows by a simple application of the SLLN.

3.8 Skorokhod's Representation Theorem

PROBLEM 3.8.1. Let[13] Q and Q_n, $n = 1, 2, \ldots$ be probability measures on $\mathbf{R}$ and suppose that $Q_n \to Q$ as $n \to \infty$ and that there exists a single bounded set that supports each Q_n. Then

$$\int_{\mathbf{R}} x\, Q_n(dx) \to \int_{\mathbf{R}} x\, Q(dx) \quad \text{as } n \to \infty.$$

Prove also that the assumption that each Q_n is supported by a single bounded set cannot be dropped.

[13]The first part of this problem is taken from Gray and Fristedt's, *A Modern Approach to Probability Theory*, op.cit.

SOLUTION. Since $Q_n \to Q$ as $n \to \infty$, by Skorokhod's Theorem we know that there exist a probability space $(\Omega, \mathcal{F}, P)$ and r.v.'s X and X_n, $n = 1, 2, \ldots$ such that $X \sim Q$, $X_n \sim Q_n$, $n = 1, 2, \ldots$ and $X_n \xrightarrow{a.s.} X$ as $n \to \infty$. Hence, to prove the statement of the problem it suffices to show that $E[X_n] \to E[X]$ as $n \to \infty$. To this purpose, let B be the single bounded set that supports Q_n, $n = 1, 2, \ldots$, i.e.: $P[X_n \in B] = 1$ for all $n = 1, 2, \ldots$ and on B we clearly have that $|X_n| \leq M$ for some constant $M \geq 0$. The statement of the problem then follows from an application of the Bounded Convergence Theorem.

To show that the assumption about each of the Q_n's being supported by a single bounded set cannot be dropped one can use the following counter-example. Let Q_n be defined by

$$Q_n(\{0\}) = 1 - \frac{1}{\sqrt{n}}, \quad Q_n(\{n\}) = \frac{1}{\sqrt{n}}$$

for $n = 1, 2, \ldots$ It is clear that the Q_n's are not supported by a single bounded subset of the real line. Now, $Q_n \to \delta_{\{0\}} \equiv Q$ as $n \to \infty$. On the other hand, we find

$$\int_{\mathbf{R}} x Q_n(dx) = \sqrt{n} \to \infty \text{ as } n \to \infty$$

but $\int_{\mathbf{R}} x Q(dx) = \int_{\mathbf{R}} x \delta_{\{0\}}(dx) = 0$.

Remark. Another lesson taught by this problem is that convergence in distribution is not enough to have convergence of moments and that, in general, further assumptions are required.

PROBLEM 41. Let Q and Q_n, $n = 1, 2, \ldots$, be probability measures on $\mathbf{R}$ and suppose that each Q_n has the interval $[0, \infty)$ as the support set. Assume also that $Q_n \to Q$ as $n \to \infty$. Prove that

$$\int_{\mathbf{R}} x\, Q(dx) \leq \liminf \int_{\mathbf{R}} x\, Q_n(dx).$$

Show also that the assumption that the Q_n's are supported by $[0, \infty)$ cannot be dropped.

SOLUTION. Since $Q_n \to Q$ as $n \to \infty$, by Skorokhod's Representation Theorem, there exist a probability space $(\Omega, \mathcal{F}, P)$ and random variables X, X_n, $n = 1, 2, \ldots$ defined on it and such that $X_n \xrightarrow{a.s.} X$ as $n \to \infty$. In addition, because of the assumption about the support set, we have also that $X_n \geq 0$ for all $n = 1, 2, \ldots$ The statement of the problem follows then by a simple application of Fatou's Lemma, i.e.:

$$E[X] \leq \liminf E[X_n]$$

which can also be written as

$$\int_{\mathbf{R}} x\, Q(dx) \leq \liminf \int_{\mathbf{R}} x\, Q_n(dx).$$

The next counter-example shows that the assumption about the support set cannot be dropped. To this purpose let Q_n be defined as follows:

$$Q_n(\{0\}) = 1 - \frac{1}{n}, Q_n(\{-n\}) = \frac{2}{3n}, Q_n(\{n\}) = \frac{1}{3n}, \quad n = 1, 2, \dots$$

It is easy to show that $Q_n \to \delta_{\{0\}}$ as $n \to \infty$ so that $\int_{\mathbf{R}} x\, Q(dx) = 0$. However, we have

$$\liminf \int_{\mathbf{R}} x\, Q_n(dx) = -\frac{1}{3}$$

which proves the point.

PROBLEM 3.8.2. Let $\{P_n : n \geq 1\}$ and $\{Q_n : n \geq 1\}$ be two sequences of probability measuress such that $P_n \to P$ and $Q_n \to Q$ as $n \to \infty$. Prove that the sequence of convolutions $\{P_n \star Q_n) : n \geq 1\}$ is such that $P_n \star Q_n \to P \star Q$ as $n \to \infty$.

SOLUTION. If one assumes that $P_n \to P$ and $Q_n \to Q$ as $n \to \infty$, the Skorokhod's Representation Theorem guarantees that there exist probability spaces $(\Omega_X, \mathcal{F}_X, P_X)$ and $(\Omega_Y, \mathcal{F}_Y, P_Y)$ and random variables X_n and X, Y_n and Y definedon Ω_X and Ω_Y, respectively, such that

$$X_n \sim P_n, \;\; Y_n \sim Q_n, \;\; X \sim P, \;\text{ and }\; Y \sim Q$$

and $X_n \overset{a.s.}{\to} X$, $Y_n \overset{a.s.}{\to} Y$ as $n \to \infty$. This implies that $(X_n, Y_n) \overset{P}{\to} (X, Y)$ and since convergence in probability implies convergence in distribution, it is also $(X_n, Y_n) \overset{D}{\to} (X, Y)$ as $n \to \infty$. Since $P_n \star Q_n$ is the probability distribution of (X_n, Y_n) and $P \star Q$ is the probability distribution of (X, Y), this fact implies that $P_n \star Q_n \to P \star Q$ as $n \to \infty$.

4 Chapter

CHARACTERISTIC FUNCTIONS AND LAPLACE TRANSFORMS

4.1 Theoretical Properties

PROBLEM 4.1.1. Show that X is symmetric about zero if and only if its characteristic function, $\phi(t)$, is real valued for all t.

SOLUTION. We start by providing few useful basic facts about characteristic functions:

1. if $\phi(t)$ is a ch.f., then we know that $\phi(-t) = \overline{\phi(t)}$;
2. (UNIQUENESS THEOREM FOR CH.F.'s:) Two d.f. F_1 and F_2 are identical iff their ch.f.'s ϕ_1 and ϕ_2 are identical.
3. X is symmetric about 0 iff X and $-X$ have the same distribution. This means that $F_X(x) = F_{-X}(x)$. There are, of course, other ways to express this fact, e.g.: the p.d.f., $f_X(x)$ is an even function on its domain of definition.
4. $\phi_X(-t) = \phi_{-X}(t)$.

The proof of the last fact is based on the fact that for any r.v. X, we have

$$\phi_X(-t) = \int_{-\infty}^{\infty} \exp\{-itx\}dF_X(x) = \int_{-\infty}^{\infty} \exp\{-itx\}d[1 - F_{-X}(-x-)]$$

as $F_X(y) = 1 - F_{-X}(-y)$ whenever $-y$ is a continuity point for $F(\cdot)$, or $F_X(y) = 1 - F_{-X}(-y-)$ in those cases where $-y$ is a discontinuity point for $F(\cdot)$. Of course, this definition is consistent as in fact $F_{-X}(-y-) = F_{-X}(-y)$ if $-y$ is a continuity point for $F(\cdot)$. With this, one finds

$$\phi_X(-t) = \int_{-\infty}^{\infty} \exp\{-itx\}d[1 - F_{-X}(-x-)]$$

and, letting $-x = y$,

$$\phi_X(-t) = \int_{\infty}^{-\infty} \exp\{ity\} d[1 - F_{-X}(y+)] =$$

$$= -\int_{\infty}^{-\infty} \exp\{ity\} d[1 - F_{-X}(y)] =$$

$$= \int_{-\infty}^{\infty} \exp\{ity\} d[1 - F_{-X}(y)] = \phi_{-X}(t).$$

This last passage can be justified using the fact that $F_{-X}(\cdot)$ being a d.f. is continuous from the right, therefore $F_{-X}(y+) = F_{-X}(y)$ $\forall y \in \mathbf{R}$.
These four facts are all we need to prove the statement of the problem. In fact,

- (if). X is symmetric about 0 then (3) implies that $F_X(y) = F_{-X}(y)$ $\forall y \in \mathbf{R}$. By (2) and (4), this implies that $\phi(t) \equiv \phi_X(t) = \phi_{-X}(t) = \phi_X(-t)$. Using (1) this gives $\overline{\phi_X(t)} = \overline{\phi(t)}$. And since for any $t \in \mathbf{R}$ $\phi(t)$ is a complex number, $\phi(t) = \overline{\phi(t)}$ is possible if and only if $\phi(\cdot)$ is a real-valued function.

- (only if). Assume that $\phi(t)$ is a real-valued c.f. $\forall t \in \mathbf{R}$. In this case, using (1) and (4), we have that:

$$\phi_X(t) \equiv \phi(t) = \overline{\phi(t)} = \phi(-t) = \phi_X(-t) = \phi_{-X}(t).$$

  Then, the uniqueness theorem (2) implies that $F_X(y) = F_{-X}(y)$ $\forall y \in \mathbf{R}$ and hence X is symmetric about the origin because of (3).

PROBLEM 4.1.2. All integrals in this problem are over $\mathbf{R}^1$. Suppose $\psi : \mathbf{R} \to [0,1]$ is continuous and satistifies $\psi(t) = \psi(-t)$, $\psi(0) = 1$, $\int \psi(s)ds < \infty$. Define g by $g(x) = (2\pi)^{-1} \int \exp\{-isx\}\psi(s)ds$. Suppose $g(x) \geq 0$ and $\int g(x)dx = 1$. Prove that ψ is a characteristic function.

SOLUTION. Let $K = \int \psi(s)ds$. Then ψ/K is a density function, as it is nonnegative and integrates to 1. Besides, $2\pi g(t)K^{-1} = \int \exp\{-isx\}\psi(s)ds = \int \exp\{isx\}g(s)ds$ is its characteristic function, where the last equality follows from the symmetry of ψ. Since $2\pi g(t)K^{-1}$ is integrable, the <u>inversion theorem</u> can be used to yield:

$$\frac{\psi(t)}{K} = (2\pi)^{-1} \int \exp\{-ist\} 2\pi g(s) K^{-1} ds,$$

or $\psi(t) = \int \exp\{-ist\}g(s)ds$ so that ψ is a characteristic function. Again, the last inequality follows from the easily verified fact that g is symmetric.

PROBLEM 4.1.3. Which of the following are characteristic functions? (Prove or disprove.)

(a) $1+t^2$;

(b) $\frac{1}{4}+\frac{\cos(t)}{2}+\frac{\sin(t)}{4t}$;

(c) $\exp\{-t^4\}$;

(d) $\frac{2}{1+\exp\{it\}}$.

SOLUTION. A description of a few facts about ch.f's which will be useful in the following is presented next.
Let F_X be a distribution function with ch.f. ϕ; then:

(1) $\phi(0) = 1$;

(2) $\mid \phi(t) \mid\leq 1\ \forall t \in \mathbf{R}$;

(3) $\phi(-t) = \overline{\phi(t)}$ ($\phi(-t) = \phi(t)$ if ϕ is a real valued function);

(4) $\phi(t)$ is uniformly continuous on the entire real line;

(5) Suppose that the real numbers $\alpha_1, \alpha_2, \ldots, \alpha_n$ satisfy the conditions:

   i. $\alpha_i \geq 0,\ i = 1, 2, \ldots, n$;
   ii. $\sum_{i=1}^{n} \alpha_i = 1$;
   iii. $\phi_1(t), \phi_2(t), \ldots, \phi_n(t)$ are c.f.s;

   then $\phi(t) = \sum_{i=1}^{n} \alpha_i\phi_i(t)$ is also a characteristic function. [For a proof of these facts, see Lukacs, Characteristic Functions, 2^{nd} edition, London 1970; Thms 2.1.1, 2.1.2 and 2.1.3]

Using the fact above we have that:

- (a) is not a ch.f. since $\phi(t) = 1+t^2$ contradicts fact (2): $\mid \phi(t) \mid=\mid 1+t^2 \mid\geq 1$ $\forall t \in \mathbf{R}$.

- (b) is a ch.f. because if we let $\alpha_i = 1/4$, $\alpha_2 = 1/2$ and $\alpha_3 = 1/4$; $\phi_1(t) = 1$ is the ch.f. for a point mass distribution ($\delta_{\{0\}}$), $\phi_2(t) = \cos(t)$ is the ch.f. for a Bernoulli(1/2) distribution with mass 1/2 at ± 1, while $\phi_3(t) = \sin(t)/t$ is the ch.f. for a uniform$(-1, 1)$ distribution. Hence, using (5) we have that:

$$\phi(t) = (1/4) \cdot \phi_1(t) + (1/2) \cdot \phi_2(t) + (1/4) \cdot \phi_3(t)$$

  is a ch.f.

- (c) is not a ch.f. as in this case we see that $\phi(t) = \exp\{-t^4\}$ is such that $\phi'(t)\mid_{t=0}= \phi''(t)\mid_{t=0}= 0$ and, therefore, $\phi(t)$ is a ch.f. for a random variable X having zero mean and zero variance. The only random variable with this characterisctics has a degenerate distribution which gives probability mass 1 to $\{0\}$. But we know that for such a distribution $\phi_X(t) = 1$ $\forall t \in \mathbf{R}$ and since $\phi(t) \neq 1 = \phi_X(t)$, it cannot be a ch.f. or a contradiction (because of the uniqueness theorem) arises. This provides an example of a function that, although it not a ch.f., satisfies (1) - (4) and makes it clear that (1) - (4) are only necessary conditions for a function on the real line to be a ch.f.

- (d) is not a ch.f. since in this particular case $\phi(t)$ fails to satisfy (2). In fact, it is easily seen that

$$\frac{2}{1+\exp\{it\}} = \frac{2}{1+\cos(t)+i\sin(t)} \to \infty$$

when $t \to \pi \pm 2k\pi$, $k = 0, 1, 2, \ldots$ and therefore:

$$\left| \frac{2}{1+\exp\{it\}} \right| > 1$$

for some t.

PROBLEM 4.1.4. Prove that

$$\phi(z) = \sqrt{1-z^2} \cdot I_{[-1,1]}(z)$$

is not a characteristic function.
Prove also that if $\psi(z)$ is another function such that

$$\lim_{z \to 0} \frac{\psi(z) - 1}{z^2} = 2$$

then ψ cannot be a characteristic function either.

SOLUTION. There is likely more than one way to show that ϕ is not a characteristic function. One way is to use the fact that if ϕ is the characteristic function of a r.v. Z and has a finite derivative of even order k at $z = 0$ then Z has a finite moment of order k and $i^k E[Z^k] = \phi^{(k)}(0)$ (see e.g. Chung, A Course in Probability Theory, 1974; Theorem 6.4.1). In our case, if we assumed that ϕ is a characteristic function, we find that

$$\phi^{(iv)}(z) = \frac{-3(1-z^2)^{3/2} - 15z^2(1-z^2)^{3/2}}{(1-z^2)^5}$$

and

$$E[X^4] = \frac{1}{i^4}\phi^{(iv)}(0) = -3$$

which is clearly impossible. Hence, ϕ is not a characteristic function. In the case of ψ, if we assumed that ψ is a characteristic function, we have an indetetermi-nate form, $\frac{0}{0}$, which we can handle using De L'Hospital's Rule twice. The fact that $\frac{\psi'(z)}{z}$ is also an indeterminate form when $z \to 0$ follows from the fact that since ψ is a real function it means that the r.v., Z, it is the characteristic function of, is symmetric and hence has zero odd moments. Then $E[Z] = \frac{1}{i}\psi'(0) = 0$. This yields

$$2 = \lim_{z\to 0} \frac{\psi(z)-1}{z^2} = \lim_{z\to 0} \frac{\psi'(z)}{2z} = \lim_{z\to 0} \frac{\psi''(z)}{2}$$

or, $\psi''(0) = 4 > 0$ which is a contradiction since $\psi''(0)$ must be negative. Hence, ψ cannot be a characteristic function.

PROBLEM 4.1.5. Let ϕ be a characteristic function. Prove that

$$\psi(z) = \frac{p}{1-(1-p)\phi(z)} \quad p \in (0,1)$$

is also a characteristic function.

SOLUTION. Let $\{Y_k : k = 1,2,\ldots\}$ be a sequence of i.i.d. random variables having characteristic function ϕ. Let also T be a random variable such that $T \sim$ Negative binomial$(1,p)$ and is stochastically independent of $\{Y_k : k = 1,2,\ldots\}$. Define the new random variable

$$Z = \sum_{k=1}^{T} Y_k.$$

Then, if ψ is the characteristic function for Z we find that

$$\begin{aligned}
\psi_Z(t) &= E[e^{itZ}] = E[e^{it\sum_{k=1}^{T} Y_k}] \\
&= E_T\left[E_{Y_1|T}[e^{itY_1}]^T \mid T\right] = E_T\left[E_{Y_1}[e^{itY_1}]^T]\right] \\
&= E_T\left[E[e^{itTY_1}]\right] = E_T[e^T E[e^{itY_1}]] \\
&= E_T[e^T \phi_{Y_1}(t)] = E_T[e^{T\log(\phi_{Y_1}(t))}] \\
&= \varphi_T(\log(\phi_{Y_1}(p))
\end{aligned}$$

where $\varphi_T(\cdot)$ is the moment generating function for T. Since $T \sim$ Negative binomial$(1,p)$, $p \in (0,1)$ the analytic form of $\varphi_T(\cdot)$ is known and it yields

$$\psi_Z(t) = \frac{p}{1-(1-p)e^{\log(\phi_{Y_1}(p))}} = \frac{p}{1-(1-p)\phi_{Y_1}(t)}$$

which proves the statement of the problem.

PROBLEM 4.1.6. Prove that the Laplace-Stielties Transform of an $\mathbf{R}^+$-valued random variable is continuous on $[0,\infty)$ and that the Laplace-Stielties Transform of an $\bar{\mathbf{R}}^+$-valued random variable is continuous on $(0,\infty)$.

SOLUTION. Assume first that X is an $\mathbf{R}^+$-valued random variable. Then its Laplace-Stielties Transform is given by

$$\varphi(u) = E[e^{-uX}] = \int_{R^+} e^{-ux} Q(dx) \quad u \geq 0.$$

When $u > 0$ and h is chosen so that $| h |< u$, we have that $u + h > 0$ and therefore $| e^{-(u+h)x} |< 1$ for all values of X. Hence, in the the following equality

$$\lim_{h\to 0} \varphi(u+h) - \varphi(u) = \lim_{h\to 0}\left[\int_{R+} e^{-(u+h)x}Q(dx) - \int_{R+} e^{-ux}Q(dx)\right]$$

it is possible to use the DCT and interchange the operations of limit and integration to get

$$\lim_{h\to 0} \varphi(u+h) - \varphi(u) = \int_{R=} \lim_{h\to 0} e^{-(u+h)x}Q(dx) - \int_{R+} e^{-ux}Q(dx)\Big] = 0.$$

This proves that $\varphi(u)$ is continuous at every $u > 0$. The next task is to prove that it is right continuous at 0. To this purpose, let $h > 0$ and consider that for each $u \geq 0$ we can write

$$\lim_{h\to 0} \varphi(u+h) - \varphi(u) = \lim_{h\to 0}\left[\int_{R+} e^{-(u+h)x}Q(dx) - \int_{R+} e^{-ux}Q(dx)\right] =$$

$$= \int_{R+} e^{-ux}[e^{-hx} - 1]Q(dx).$$

In this case $| e^{-ux} |\geq 1$ and $| e^{-hx} - 1 |\geq 1$ so that by the DCT we can interchange once more the operations of limit and integration to find

$$\lim_{h\to 0} \varphi(u+h) - \varphi(u) = \int_{R+} e^{-ux}[\lim_{h\searrow 0}(e^{-hx} - 1)Q(dx) = \int_{R+} e^{-ux}\cdot 0\cdot Q(dx) = 0.$$

In the case of an $\bar{\mathbf{R}}^+$-valued random variable if we consider $u > 0$ we have

$$\lim_{h\to 0} \varphi(u+h) - \varphi(u) = \lim_{h\to 0}\left[\int_{\bar{R}+} e^{-(u+h)x}Q(dx) - \int_{\bar{R}+} e^{-ux}Q(dx)\right] =$$

$$= \lim_{h\to 0}\int_{R+} e^{-(u+h)x}Q(dx) + e^{-(u+h)\infty}Q(\{\infty\}) - \int_{R+} e^{-ux}Q(dx) - e^{-u\infty}Q(\{\infty\}).$$

Using the DCT since $| e^{-(u+h)x} |< 2$ for sufficiently small h we can rewrite the expression above as

$$\int_{R+} \lim_{h\to 0} e^{-(u+h)x}Q(dx) - \int_{R+} e^{-ux}Q(dx) + \lim_{h\to 0} e^{-(u+h)\infty}Q(\{\infty\}) - e^{-u\infty}Q(\{\infty\})$$

$$= \int_{R+} e^{-ux}Q(dx) - \int_{R+} e^{-ux}Q(dx) + e^{-u\infty}Q(\{\infty\}) - e^{-u\infty}Q(\{\infty\}) = 0.$$

Thus we have established that $\varphi(u)$ is continuous on $(0, \infty)$.
If $u = 0$, replacing u by 0 in the last expression above, we have that

$$\lim_{h\searrow 0} \varphi(0+h) - \varphi(0) = \lim_{h\searrow 0}\left[\int_{\bar{R}+} e^{-hx}Q(dx) - \int_{\bar{R}+} e^{-0x}Q(dx)\right] =$$

$$= \lim_{h \searrow 0} \int_{R+} e^{-hx} Q(dx) + \lim_{h \searrow 0} e^{-h\infty} Q(\{\infty\}) - \int_{\bar{R}+} e^{-0x} Q(dx).$$

Interchanging the limit and integral signs we find that

$$\lim_{h \searrow 0} \int_{R+} e^{-hx} Q(dx) = \int_{R+} \lim_{h \searrow 0} e^{-hx} Q(dx) = \int_{R+} e^{-0} Q(dx) = Q(\mathbf{R}^+) =$$
$$= 1 - Q(\{\infty\})$$

and hence

$$\lim_{h \searrow 0} \varphi(0+h) - \varphi(0) = 1 - Q(\{\infty\}) + 0 - Q(\bar{\mathbf{R}}^+) = Q(\{\infty\})$$

since $Q(\bar{\mathbf{R}}^+) = 1$. As X is an $\bar{\mathbf{R}}^+$-valued random variable it must be $Q(\{\infty\}) \neq 0$ and thus $\varphi(u)$ cannot be continuous at 0.

PROBLEM 4.1.7. Let F be a distribution function and let ϕ be the corresponding characteristic function. Then, if F is absolutely continuous,

$$\lim_{|t| \to \infty} \phi(t) = 0.$$

In the general case, if the absolutely continuous part of F does not vanish, then

$$\limsup_{t \to \infty} |\phi(t)| < 1.$$

SOLUTION. Let F be absolutely continuous with density f and we start by assuming that f is a step function, i.e.:

$$f(x) = \sum_{j=1}^{k} \alpha_j \cdot I_{A_j}(x)$$

where $\{A_j = (a_j, b_j] : j = 1, 2, \ldots, k\}$ is a partition of $\mathbf{R}$. Then

$$\phi(t) = \int_{-\infty}^{\infty} e^{itx} f(x)\, dx = \sum_{j=1}^{k} \alpha_j \int_{a_j}^{b_j} e^{itx}\, dx = \sum_{j=1}^{k} \frac{e^{itb_j} - e^{ita_j}}{it}.$$

Clearly $\phi(t) \to 0$ as $t \to \infty$ as the sum is finite. In addition, as the set of step functions is dense in L^1, for every general density, f, and for every positive real number, ϵ, it is possible to find a step function g_ϵ such that

$$\int |f(x) - g_\epsilon(x)|\, dx \leq \epsilon.$$

This yields:

$$\epsilon \geq \int |f(x) - g_\epsilon(x)|\, dx \geq \int |e^{itx}(f(x) - g_\epsilon(x))|\, dx$$
$$\geq |\phi(t) - \int e^{itx} g_\epsilon(x)\, dx| \geq |\phi(t)| - |\int e^{itx} g_\epsilon(x)\, dx|.$$

From the last inequality, it is easily seen that

$$\limsup_{t\to\infty} |\phi(t)| \leq \epsilon$$

and, since ϵ is arbitrary, it is also

$$\lim_{t\to\infty} |\phi(t)| = 0.$$

In the general case, if we assume that

$$F = \gamma_1 F_1 + \gamma_2 F_2$$

where F_1 is the absolutely continuous part of F and $\gamma_1, \gamma_2 > 0$, $\gamma_1 + \gamma_2 = 1$, we find

$$\phi(t) = \gamma_1\phi_1(t) + \gamma_2\phi_2(t)$$

where ϕ_1 and ϕ_2 are the characteristic functions corresponding to F_1 and F_2, respectively. Thus, using the fact that for an absolutely continuous distribution function we have $\lim t \to \infty |\phi(t)| = 0$, we have

$$\limsup_{t\to\infty} |\phi(t)| \leq \gamma_1 \limsup_{t\to\infty} |\phi_1(t)| + \gamma_2 \limsup_{t\to\infty} |\phi_2(t)| \leq \gamma_2 < 1.$$

PROBLEM 4.1.8. Assume that ϕ is a ch.f. Show that $\mathrm{Re}\,\phi$ and $|\phi|^2$ are also ch.f.'s.

SOLUTION. We state a couple of simple facts about ch.f.'s that will be useful towards proving the statement of the problem.

Fact A: If $F_1, F_2, \ldots F_n$ are distribution functions having ch.f. $\phi_1, \phi_2, \ldots, \phi_n$ and $\{\alpha_i : i = 1, 2, \ldots, n\}$ are real numbers such that $\alpha_i \geq 0$ $i = 1, 2, \ldots, n$ and $\sum_{i=1}^n \alpha_i = 1$, then $\sum_{i=1}^n \alpha_i F_i$ has ch.f. $\sum_{i=1}^n \alpha_i \phi_i$.

Fact B: If X and Y are independent r.v.'s having ch.f.'s ϕ_X and ϕ_Y, resptectively then $X + Y$ has ch.f. $\phi_X \cdot \phi_Y$.

Since

$$\mathrm{Re}\,\phi = \frac{\phi + \bar{\phi}}{2} \quad \text{and} \quad |\phi|^2 = \phi \cdot \bar{\phi},$$

all one has to show is that $\frac{\phi+\bar{\phi}}{2}$ and $\phi \cdot \bar{\phi}$ are ch.f.'s.
We know that if X has ch.f. $\phi(t)$, then $_X$ has ch.f. given by $\phi(-t) = \bar{\phi(t)}$. This together with Fact A proves that $\mathrm{Re}\,\phi$ is a ch.f.
In addition, if we let Z and Y be two independent r.v.'s with the same distribution as X and $-X$, respectively, we can use Fact B and prove that $|\phi|^2$ is also a ch.f.

PROBLEM 4.1.9. Show that if $\phi(t)$ is a ch.f., $\exp\{\alpha\cdot(\phi(t)-1)\}$ is also a ch.f. Then, show that

$$\chi(t)=\exp\{\alpha(\cos t-1)-\beta t^2-\gamma|t|\}$$

where $\alpha,\beta,\gamma>0$ is a ch.f. as well.

SOLUTION. Since $\phi_1(t)\equiv 1$ is a ch.f. (for the random variable $\delta_{\{0\}}$) and ϕ is also a ch. f., so is their convex combination

$$\left(1-\frac{\alpha}{n}\right)\cdot\phi_1+\frac{\alpha}{n}\cdot\phi=1+\frac{\alpha}{n}[\phi-1].$$

Now, $\psi_n(t)=(1+\alpha/n[\phi(t)-1])^n$ is the ch.f. of $X_1+X_2+\ldots+X_n$, where the X_i's are i.i.d. r.v. and X_1 has ch.f. $1+\frac{\alpha}{n}(\phi-1)$. Since

$$\lim_n \psi_n(t)=\lim_n(1+\alpha/n[\phi(t)-1])^n=\exp\{\alpha(\phi(t)-1)\}$$

and since the limit is clearly continuous at 0, it follows that $\exp\{\alpha(\phi(t)-1)\}$ must be a ch.f. as a consequence of the continuity theorem.

Since $\cos t$ is a ch.f. and, more precisely the ch.f. of the random variable

$$X=\begin{cases}1 & \text{with probability } 1/2\\ -1 & \text{with probability } 1/2,\end{cases}$$

it follows from the previous part of the problem that $\exp\{\alpha(\cos t-1)\}$ is a c.f. In addition, $\exp\{-\beta t^2\}$ is the ch.f. of a $N(0,2\beta)$ random variable and $\exp\{\gamma|t|\}$ is the ch.f. of a Cauchy(γ) random variable. Then, as $\chi(t)$ is the product of ch.f.'s, it is itself a ch.f.

PROBLEM 4.1.10. A characteristic function ϕ is said to be infinitely divisible, if for every positive integer, n, it is the n-th power of some characteristic function ψ_n, i.e. $\phi=[\psi_n]^n$. Such function is uniquely determined by ϕ taking the principal branch as the n-th root.

Prove that if ϕ is an infinite divisible ch.f. then it has no real zeros.

Prove that $|\phi|^2$ is also an infinitelely divisible ch.f.

SOLUTION. By definition, as ϕ is an infinite divisible ch.f., for any integer n we can find a ch.f. ψ_n such that

$$\phi(t)=|\psi_n(t)|^n\ \Rightarrow \psi_n(t)=[\phi(t)]^{1/n}.$$

Let

$$h(t)\equiv\lim_n[\psi_n(t)]^2=\lim_n[\phi(t)]^{2/n}=\begin{cases}0 & \text{if } \phi(t)=0;\\ 1 & \text{if } \phi(t)\neq 0.\end{cases}$$

It follows wasily that $h(0)=1$. In addition, since ϕ is a ch.f. it is continuous in a neighborhood of the origin and $\phi(0)=1$. This means that it is possible to

find a real number δ such that $h(t) = 1\ \forall t \in B(0,\delta)$. In particular, if follows that $h(t)$ is continuous at the origin. Now, $h(t)$ is a limit of ch.f.'s (we saw in a previous problem that if ψ_n is a ch.f., so is $|\psi_n|^2$) and it is continuous at 0, therefore it is itself a ch.f. and, hence, it is (uniformly) continuous on the entire real line. This can happen iff $h(t) \equiv 1$ and thus it must be $\phi(t) \neq 0\ \forall t \in \mathbf{R}$ which is the conclusion we were trying to establish.
We just mention the fact that the result above provides only a necessary condition for a ch.f. to be infinitely divisible. To this purpose, see E. Lukacs, Characteristic Functions, 2nd Ed., 1970; p. 109.
In order to prove the second statement of the problem, we introduce the following

Lemma. *Let X and Y be infinitely divisible and independent r.v.'s on some probability space $(\Omega, \mathcal{F}, P)$. Then, $X+Y$ is an infinite divisible r.v.*

Proof. Assume that X and Y are as in the text of the Lemma and have ch.f.'s $\phi_X(t)$ and $\psi_Y(t)$, respectively. Then, $Z = X+Y$ has ch.f. $\chi_Z(t) = \phi_X(t)\cdot\psi_Y(t)$ since, by assumption, X and Y are independent. Using the assumption about X and Y being infinitely divisible, for any $n \geq 1$ it is possible to find ch.f.'s $\phi_n(t)$ and $\psi_n(t)$ such that

$$\phi_X(t) = [\phi_n(t)]^n \text{ and } \psi_Y(t) = [\psi_n(t)]^n,$$

respectively. Thus, for any $n \geq 1$ the ch.f. of Z can be written as

$$\chi_Z(t) = \phi_X(t)\cdot\psi_Y(t) = [\phi_n(t)]^n \cdot [\psi_n(t)]^n. = [\phi_n(t)\cdot\psi_n(t)]^n.$$

As the product of ch.f.'s is a ch.f., we have proved that $Z = X+Y$ is also an infinitely divisible r.v. □

Now, let X be a r.v. on some probability space $(\Omega, \mathcal{F}, P)$ and with ch.f. $\phi(t)$. Let Y be another r.v. on the same probability space, independent of X and with the same distribution. Putting $Z = X + (-Y)$, we know from the Lemma above that Z is an infinitely divisible r.v. and so its ch.f. $\chi_Z(t)$ is an infinite divisible one. But, we have also

$$\chi_Z(t) = \phi(t)\cdot\phi(-t) = |\phi(t)|^2.$$

This two last facts together imply the second statement of the problem.

PROBLEM 4.1.11. Show that if X is a $\mathbf{Z}$-valued random variable defined on a probability space $(\Omega, \mathcal{F}, Q)$ and $\phi_X(t)$ is its ch.f., then

$$P[X = k] = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itk}\phi_X(t)\,dt \quad \forall k \in \mathbf{Z}.$$

Show that more in general, if P is a probability measure on some sample space $(\Omega, \mathcal{F})$,

$$P[\{k\}] = \lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T} e^{-itk}\phi(t)\,dt.$$

SOLUTION. It is easily seen that $\phi_X(t)$ is 2π-periodic. In fact,

$$\phi_X(t+2\pi) = \sum_{n\in\mathbf{Z}} e^{2i\pi n} e^{itn} Q(n) = \sum_{n\in\mathbf{Z}} e^{itn} Q(n) = \phi_X(t)$$

since $e^{2i\pi n} = \cos(2\pi n) + i\sin(2\pi n) = 1 + i0 = 1$. Now, using the definition of ch.f. for a **Z**-valued random variable, we can write

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itk}\phi_X(t)\,dt = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itk}\Big(\sum_{n\in\mathbf{Z}} e^{itn} Q(\{n\})\Big)\,dt$$

and, using Fubini's Theorem,

$$= \sum_{n\in\mathbf{Z}}\Big(\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{it(n-k)}\,dt\Big) Q(\{n\}),$$

in addition, since

$$e^{it(n-k)} = \cos[t(n-k)] + i\sin[t(n-k)] = \begin{cases} 0 & \text{if } n \neq k \\ 1 & \text{if } n = k \end{cases},$$

we get

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-itk}\phi_X(t)\,dt = \frac{1}{2\pi}\int_{-\pi}^{\pi} I_{\{n=k\}} e^{it(n-k)}\,dt Q(\{k\}) = \frac{2\pi}{2\pi} Q(\{k\}) = Q(\{k\})$$

as we were supposed to show.
The second case is almost similar, i.e.

$$\frac{1}{2T}\int_{-T}^{T} e^{-itk}\phi(t)\,dt = \frac{1}{2T}\int_{-T}^{T} e^{-itk}\int_{-\infty}^{\infty} e^{-itx} P(dx)\,dt$$

using Fubini's Theorem to interchange the order of integration we get

$$\begin{aligned} &= \int_{-\infty}^{\infty} \frac{1}{2T}\int_{-T}^{T} e^{it(x-k)} dt\, P(dx) \\ &= \int_{-\infty}^{\infty} \frac{1}{2T}\Big[\int_{-T}^{T}\cos[t(x-k)]dt + i\int_{-T}^{T}\sin[t(x-k)]dt\Big] P(dx) \end{aligned}$$

and using the fact that sin is an odd function, we can simplify the last formula and write

$$\frac{1}{2T}\int_{-T}^{T} e^{-itk}\phi(t)\,dt = \int_{-\infty}^{\infty}\frac{1}{2T}\Big[\int_{-T}^{T}\cos[t(x-k)]dt\Big] P(dx)$$

and since

$$|\frac{1}{2T}[\int_{-T}^{T}\cos[t(x-k)]dt| \leq 1 \quad\text{and}\quad \int_{-T}^{T} 1P(dx) = 1 < \infty$$

we can use the Bounded Convergence Theorem for functions and write

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-itk}\phi(t)\,dt = \lim_{T\to\infty} \frac{1}{2T} \left[\int_{-T}^{T} \cos[t(x-k)]dt\right] P(dx)$$
$$= \int_{-\infty}^{\infty} \left(\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \cos[t(x-k)]dt\right) P(dx)$$

As

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} \cos[t(x-k)]dt = \begin{cases} 0 & \text{if } x \neq k \\ 1 & \text{if } x = k \end{cases},$$

we have

$$\lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{-itk}\phi(t)\,dt = \int_{-\infty}^{\infty} I_{\{x=k\}} P(dx) = P(\{k\})$$

as we were supposed to prove.

4.2 Applications

PROBLEM 4.2.1. Let $\mathbf{X} = (X_1, X_2, X_3)^T$, $\mathbf{t} = (t_1, t_2, t_3)^T$ and

$$\phi_{\mathbf{X}}(\mathbf{t}) = \exp\left\{-\frac{1}{4}\sum_{i=1}^{3} t_i^2\right\} \cdot \left[\sum_{i=1}^{3} \exp\left\{\frac{-t_i^2}{4}\right\} - 2\right].$$

Find

(a) the marginal distribution of X_3;

(b) the bivariate distribution of $\mathbf{X}^{(1)} = (X_1, X_2)^T$;

(c) $E[\mathbf{X}^{(1)}]$ and $\mathrm{Cov}[\mathbf{X}^{(1)}]$;

Finally, determine whether X_1 and X_2 are independent.

SOLUTION[1]. The ch.f. of the marginal distribution of X_3 can be obtained from $\phi_{\mathbf{X}}(\mathbf{t})$ setting $t_1 = t_2 = 0$. This gives

$$\phi_{X_3}(t_3) = \exp\left\{\frac{-t_3^2}{4}\right\} \cdot \exp\left\{\frac{-t_3^2}{4}\right\} = \exp\left\{\frac{-t_3^2}{2}\right\}$$

and, since the ch.f. determines the distribution function uniquely, we have established that $X_3 \sim N(0,1)$.
The ch.f. for $\mathbf{X}^{(1)}$ is computed setting $t_3 = 0$ in the expression for $\phi_{\mathbf{X}}(\mathbf{t})$, i.e.

$$\phi_{\mathbf{X}^{(1)}}(\mathbf{t}^{(1)}) = \exp\left\{\frac{-(t_1^2+t_2^2)}{4}\right\} \cdot \left[\exp\left\{\frac{-t_1^2}{4}\right] + \exp\left\{\frac{-t_2^2}{4}\right\} - 1\right\}.$$

[1]The solution to this problem was provided by Panagiotis Tsiamyrtzis.

Using the inversion formula, one finds

$$\begin{aligned} f_{\mathbf{X}^{(1)}}(x_1, x_2) &= \frac{1}{(2\pi)^2} \lim_{T\to\infty} \int_{-T}^{T}\int_{-T}^{T} \exp\{-i(t_1x_1 + t_2x_2\}\phi_{\mathbf{X}^{(1)}}(\mathbf{t}^{(1)})\, dt_1 dt_2 \\ &= \frac{1}{4\pi^2} \lim_{T\to\infty} \int_{-T}^{T}\int_{-T}^{T} \exp\{-i(t_1x_1 + t_2x_2\} \\ &\qquad \cdot \exp\left\{-\frac{t_1^2+t_2^2}{4}\right\}\cdot[\exp\{-t_1^2/4\} + \exp\{-t_2^2/4\} - 1]\, dt_1 dt_2 \\ &= I_1 + I_2 - I_3 \end{aligned}$$

where

$$I_1 = \frac{1}{4\pi^2}\cdot \lim_{T\to\infty} \int_{-T}^{T}\int_{-T}^{T} \exp\left\{-it_1x_1 - it_2x_2 - \frac{2t_1^2}{4} - \frac{t_2^2}{4}\right\} dt_1 dt_2,$$

$$I_2 = \frac{1}{4\pi^2}\cdot \lim_{T\to\infty} \int_{-T}^{T}\int_{-T}^{T} \exp\left\{-it_1x_1 - it_2x_2 - \frac{t_1^2}{4} - \frac{2t_2^2}{4}\right\} dt_1 dt_2,$$

and

$$I_3 = -\frac{1}{4\pi^2}\cdot \lim_{T\to\infty} \int_{-T}^{T}\int_{-T}^{T} \exp\left\{-it_1x_1 - it_2x_2 - \frac{t_1^2}{4} - \frac{t_2^2}{4}\right\} dt_1 dt_2.$$

To evaluate I_1 we observe that its expression can be rewritten as

$$I_1 = \frac{1}{4\pi^2}\cdot \lim_{T\to\infty} \int_{-T}^{T} \exp\left\{-it_1x_1 - \frac{t_1^2}{2}\right\}\cdot\left[\int_{-T}^{T}\int_{-T}^{T} \exp\left\{-it_2x_2 - \frac{t_2^2}{4}\right\}dt_2\right]dt_1$$

and that

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-irx-r^2/2}dr = \frac{1}{\sqrt{2\pi}}\cdot e^{-x^2/2}$$

using the inversion formula for normal distributions. Now, using a change of variables: $s/\sqrt{2} = t$, we find that

$$\begin{aligned} &\frac{1}{2\pi}\lim_{T\to\infty}\int_{-T}^{T} e^{-isx-s^2/4}ds \\ &\qquad = \frac{1}{2\pi}\lim_{T\to\infty}\int_{-T/\sqrt{2}}^{T/\sqrt{2}} e^{-it(\sqrt{2}x)-t^2/2}\cdot\sqrt{2}dt = \frac{\sqrt{2}}{\sqrt{2\pi}}e^{-(\sqrt{2}x)^2/2} \end{aligned}$$

and thus

$$\frac{1}{2\pi}\lim_{T\to\infty}\int_{-T}^{T} e^{-isx-s^2/4}ds = \frac{1}{\sqrt{\pi}}\cdot\exp\{-x^2\}.$$

This allows us to write

$$\begin{aligned} I_1 &= \frac{1}{\sqrt{\pi}}\cdot\exp\{-x^2\}\cdot\frac{1}{2\pi}\lim_{T\to\infty}\int_{-T}^{T} e^{it_1x_1-t_1^2/2}dt_1 \\ &= \frac{1}{\sqrt{\pi}}\cdot\exp\{-x_2^2\}\cdot\frac{1}{\sqrt{2\pi}}\exp\{-x_1^2/2\}. \end{aligned}$$

We have therefore established that

$$I_1 = \frac{1}{\sqrt{2}\pi}\exp\{-x_2^2 - x_1^2/3\}.$$

Similarly, one proves that

$$I_2 = \frac{1}{\sqrt{2}\pi}\exp\{-x_1^2 - x_2^2/2\}$$

and

$$I_3 = \frac{1}{\pi}\exp\{-x_1^2 - x_2^2\}.$$

Combining these results together, we find that

$$f_{\mathbf{X}^{(1)}}(x_1, x_2) = \frac{1}{\sqrt{2}\pi}\left[\exp\left\{-\frac{x_1^2}{2} - x_2^2\right\} + \exp\left\{-x_1^2 - \frac{x_2^2}{2}\right\} - \sqrt{2}\exp\{-x_1^2 - x_2^2\}\right].$$

From $\phi_{\mathbf{X}}(\mathbf{t})$ and letting $t_2 = t_3 = 0$ and $t_1 = t_3 = 0$ one finds that $X_1 \sim N[0,1]$ and $X_2 \sim N[0,1]$, respectively. We have therefore

$$E[\mathbf{X}^{(1)}] = E\left[\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}\right] = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and

$$Cov[\mathbf{X}^{(1)}] = \begin{pmatrix} Var[X_1] & Cov[X_1, X_2] \\ Cov[X_2, X_1] & Var[X_2] \end{pmatrix} = \begin{pmatrix} 1 & E[X_1 \cdot X_2] \\ E[X_2 \cdot X_1] & 1 \end{pmatrix}.$$

$$\begin{aligned} E[X_1 \cdot X_2] = &\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 \frac{1}{\sqrt{2}\pi} e^{-x_1^2/2 - x_2^2} dx_1 dx_2 \\ &+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 \frac{1}{\sqrt{2}\pi} e^{-x_1^2 - x_2^2/2} dx_1 dx_2 \\ &- \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 \frac{1}{\sqrt{2}\pi} e^{-x_1^2 - x_2^2} dx_1 dx_2 = 0, \end{aligned}$$

If X_1 and X_2 were independent we should find that

$$\phi_{\mathbf{X}^{(1)}}(\mathbf{t}^{(1)}) = \phi_{X_1}(t_1) \cdot \phi_{X_2}(t_2) \ \forall (t_1, t_2) \in \mathbf{R}^2,$$

but

$$\phi_{\mathbf{X}^{(1)}}(\mathbf{t}^{(1)}) = \exp\left\{-\frac{t_1^2 + t_2^2}{4}\right\} \cdot \left[\exp\left\{-\frac{t_1^2}{4}\right\} + \exp\left\{-\frac{t_2^2}{4}\right\} - 1\right]$$

and, since $X_1, X_2 \sim N[0,1]$, it is easily seen that

$$\phi_{X_1}(t_1) \cdot \phi_{X_2}(t_2) = \exp\left\{-\frac{t_1^2 + t_2^2}{2}\right\}.$$

It now suffices to verify that the two expressions differ for at least one choice of (t_1, t_2) such as $(1,1)$. Hence, X_1, X_2 are not independent.
This problem shows that $Cov(X_1, X_2) = 0$ and X_1, X_2 both $N(0,1)$ random variables is not enough to have independence. A jointly normal distribution is necessary.

PROBLEM 4.2.2. Provide an example of two real-valued r.v.'s X and Y on the same probability space $(\Omega, \mathcal{F}, P)$ such that each random variable has a nondegenerate Gaussian distribution, but (X, Y) does not have a two-dimensional Gaussian distribution.

SOLUTION. Let X be $N(0,1)$ and $Y = \delta X$ where δ is a r.v. on some probability space $(\Omega, \mathcal{F}, P)$ which takes values ± 1 with probability $1/2$ and is independent from X. Then, the ch.f. of Y is given by

$$\phi_Y(t) = E[e^{it\delta X}] = \frac{1}{2} \cdot \left[E[e^{itX}] + E[e^{-itX}]\right] = e^{-t^2/2}$$

since X is $N(0,1)$. Thus, we have established that Y is also $N(0,1)$. One finds also that

$$E[XY] = E[\delta X^2] = \int_{\delta=1} \delta X^2 dP + \int_{\delta=-1} \delta X^2 dP = \frac{1}{2} \cdot [E[X^2] - E[X^2]] = 0.$$

Therefore X and Y are uncorrelated and, hence if (X, Y) had a joint normal distribution, then X, Y would be independent. But, the joint ch.f. turns out to be

$$\begin{aligned}
\phi_{X,Y}(t_1, t_2) &= E[e^{i(t_1 X + t_2 \delta X}] \\
&= \frac{1}{2} E[e^{i(t_1+t_2)X}] + \frac{1}{2} E[e^{i(t_1-t_2)X}] \\
&= \frac{1}{2}[e^{-(t_1+t_2)^2/2} + e^{-(t_1-t_2)^2/2}]
\end{aligned}$$

since X is $N(0,1)$. It is easily seen that this expression is not equal to

$$e^{-(t_1^2+t_2^2)/2}$$

for all $t_1, t_2 \in \mathbf{R}$.

PROBLEM 4.2.3. Show that for the scalar-valued AR(1) time series with nondegenerate distribution, i.e. the stochastic process defined as

$$X_n = \rho X_{n-1} + \epsilon_n$$

where $\epsilon_1, \epsilon_2, \ldots$ are i.i.d. $N(0, \tau^2)$ with $\tau^2 > 0$,

(a) When $|\rho| < 1$, the unique invariant distribution is given by

$$X_n \sim N(0, \tau^2/(1-\rho^2)).$$

(b) When $|\rho| \geq 1$, there is no invariant probability distribution.

SOLUTION. Using the definition of an AR(1) process, i.e. $X_n = \rho X_{n-1} + \epsilon_n$ where $\epsilon_i \sim N(0, \tau^2)$, $i = 1, 2, \ldots$ we can write the ch.f. of X_n, $\psi_{X_n}(t)$, as

$$\psi_{X_n}(t) = \psi_{\epsilon_n}(t) + \psi_{X_{n-1}}(\rho t).$$

If an invariant distribution exists, it must be $psi_{X_n}(t) = \psi_{X_{n-1}}(t) = \psi(t)$ and, thus, one can write

$$\psi(t) = \exp\{-\tau^2 t^2/2\} \cdot \psi(\rho t).$$

Using this expression recursively, it is possible to rewrite the last expression as

$$\psi(t) = \exp\{-\tau^2 t^2/2[1 + \rho^2 + \rho^4 + \ldots + \rho^{2k}]\} \cdot \psi(\rho^k t).$$

If $|\rho| < 1$, when $k \to \infty$ one gets

$$\psi(t) = \exp\left\{\frac{-\tau^2 t^2}{2(1 - \rho^2)}\right\}$$

which is the ch.f. for a $N(0, \tau^2/(1 - \rho^2))$ random variable. The uniqueness part of statement (a) follows from the uniqueness property of ch.f.'s.

If, on the contrary, $|\rho| > 1$, using the same approach one finds that when $k \to \infty$, we have:

$$\psi(0) = 1; \text{ and } \psi(t) = \exp\{(-\tau^2 t^2/2) \cdot \sum_{i=1}^{k} \rho^{2i} \psi(\rho^k t) \to 0$$

for any $t > 0$. This proves that when $|\rho| \geq 1$ the ch.f. $\psi(\cdot)$ is not continuous at 0 and, therefore, there cannot be any invariant distribution for the problem at hand.

PROBLEM 4.2.4. For a stationary, scalar-valued AR(1) stochastic process

$$X_n = \rho X_{n-1} + \epsilon_n, \quad \epsilon_n \sim i.i.d.\ N(0, \tau^2),$$

$n = 1, 2, \ldots, \tau^2 > 0$, show that, for any initial distribution,

(a) the marginal distribution of X_n converges to the invariant distribution $N(0, \sigma^2)$ with $\sigma^2 = \frac{\tau^2}{1-\rho^2}.$;

(b) $\sqrt{n}\bar{X}_n \xrightarrow{D} N(0, \sigma^2 \cdot \frac{1+\rho}{1-\rho})$.

SOLUTION. It is easy to show that the AR(1) process can be written in the form

$$X_n = \rho^n \cdot X_0 + \sum_{i=1}^{n} \rho^{n-i} \epsilon_i \equiv \rho^n \cdot X_0 + \nu_n$$

where $\nu_n \sim N\left(0, \tau^2\left(\frac{1-\rho^{2n}}{1-\rho^2}\right)\right)$ and $X_0 \sim f_0$ where F_0 represents a generic distribution.
Using ch.f's, the relationship above can be expressed in the form

$$\psi_{X_n}(t) = \psi_{\nu_n}(t) \cdot \psi_{F_0}(\rho^n t) \to \exp\left\{-\frac{1}{2}\frac{\tau^2}{1-\rho^2} \cdot t^2\right\} \cdot \psi_{F_0}(0)$$

when $n \to \infty$. Since $\psi_{F_0}(0) = 1$ for any distribution F_0, the first statement of the problem is proved.
To prove the second statement, one should note that it is possible to write $\sum_{i=1}^n X_i$ as

$$\sum_{i=1}^n X_i = \frac{1-\rho^n}{1-\rho} \cdot X_0 + \left[\epsilon_n + (1+\rho)\epsilon_{n-1} + \frac{1-\rho^3}{1-\rho}\epsilon_{n-2} + \ldots + \frac{1-\rho^{n-1}}{1-\rho}\epsilon_2 + \frac{1-\rho^n}{1-\rho}\epsilon_1\right],$$

or, equivalently,

$$\sum_{i=1}^n X_i = \frac{1-\rho^n}{1-\rho} \cdot X_0 + \mu_n$$

with

$$\mu_n = \epsilon_n + (1+\rho)\epsilon_{n-1} + \frac{1-\rho^3}{1-\rho}\epsilon_{n-2} + \ldots + \frac{1-\rho^{n-1}}{1-\rho}\epsilon_2 + \frac{1-\rho^n}{1-\rho}\epsilon_1.$$

From this last expression and some algebraic manipulations, one can show that

$$\mu_n \sim N\left[0, \frac{\tau^2}{1-\rho)^2}\left((n+1) - 2 \cdot \frac{1-\rho^{n+1}}{1-\rho} + \frac{1-\rho^{2n-2}}{1-\rho}\right)\right].$$

Now, letting

$$\alpha_n = \frac{1-\rho^n}{1-\rho} \text{ and } \beta_n = -2 \cdot \frac{1-\rho^{n+1}}{1-\rho} + \frac{1-\rho^{2n-2}}{1-\rho}$$

it is possible to write the ch.f. for $\sqrt{n}\bar{X}_n$ as

$$\psi_{\sqrt{n}\bar{X}_n}(t) = \psi_{X_0}\left(\frac{\alpha_n \cdot t}{\sqrt{n}}\right) \cdot \exp\left\{\frac{-\tau^2 \cdot t^2}{2(1-\rho)^2}\left[\frac{n+1}{n} + \frac{\beta_n}{n}\right]\right\}.$$

When $n \to \infty$, it is easy to check that

$$\psi_{\sqrt{n}\bar{X}_n}(t) \to \psi_{X_0}(0) \cdot \exp\left\{\frac{-\tau^2 \cdot t^2}{2(1-\rho)^2}[1+0]\right\}$$

which is recognizable as the second statement of the problem after replacing τ^2 with its expression $\sigma^2 \cdot (1-\rho^2)$.

PROBLEM 4.2.5. Show[1] that if $X_1, X_2, \ldots X_m$ and Y are scalar zero-mean random variables, a necessary and sufficient condition for the regression of Y on $X_1, X_2, \ldots X_m$ to be linear, i.e.: $E[Y \mid \mathbf{X} = \mathbf{x}] = \sum_{j=1}^m \alpha_j x_j$, is that there exist constants α_j, $j = 1, 2, \ldots, m$ such that

$$\frac{\partial}{\partial s}\psi_{Y,X_1,\ldots,X_m}(s, t_1, \ldots, t_m)\bigg|_{s=0} = \sum_{j=1}^m \alpha_j \cdot \frac{\partial}{\partial t_j}\psi_{Y,X_1,\ldots X_m}(0, t_1, \ldots, t_m),$$

where $\psi_{Y,X_1,\ldots X_m}(s, t_1, \ldots, t_m)$ is the characteristic function (ch.f.) of Y, X_1, $\ldots, X_m$.

SOLUTION. By definition,

$$\psi_{Y,X_1,\ldots X_m}(s, t_1, \ldots, t_m) = \underbrace{\int\int\cdots\int}_{(m+1)-\text{times}} \exp\left\{isy + i\sum_{j=1}^m t_j x_j\right\} dF(y, x_1, \ldots, x_m).$$

Differentiating both sides partially with respect to s and interchanging the order of differentiation and integration on the right-hand side, one finds

$$\frac{\partial}{\partial s}\psi_{Y,X_1,\ldots,X_m}(s, t_1, \ldots, t_m) = \underbrace{\int\int\cdots\int}_{(m+1)-\text{times}} iy \cdot \exp\left\{isy + i\sum_{j=1}^m t_j x_j\right\} dF(y, x_1, \ldots, x_m),$$

which, can be rewritten as,

$$\frac{\partial}{\partial s}\psi_{Y,X_1,\ldots,X_m}(s, t_1, \ldots, t_m) = \underbrace{\int\int\cdots\int}_{(m+1)-\text{times}} iy \cdot \exp\left\{isy + i\sum_{j=1}^m t_j x_j\right\} F(dy \mid \mathbf{x}) \cdot F_{\mathbf{X}}(d\mathbf{x}),$$

where $F_{\mathbf{X}}(\mathbf{x})$ is the marginal distribution of $\mathbf{X} = (X_1, \ldots, X_m)$. Letting $s = 0$ and integrating the right-hand term with respect to y, it is easily recognized that the last expression is equivalent to

$$\frac{\partial}{\partial s}\psi_{Y,X_1,\ldots,X_m}(s, t_1, \ldots, t_m)\bigg|_{s=0} = \underbrace{\int\int\cdots\int}_{m-\text{times}} iE[y \mid \mathbf{X} = \mathbf{x}] \cdot \exp\left\{i\sum_{i=j}^m t_j x_j\right\} F_{\mathbf{X}}(d\mathbf{x}).$$

[1]This problem is a straightforward extension of C.R. Rao, Note on a Problem by Ragnar Frisch, Econometrica 15 (1947), pp. 245-9.

Now, using the assumption that $E[Y \mid \mathbf{X} = \mathbf{x}] = \sum_{j=1}^{m} \alpha_j x_j$, it is possible to rewrite the last expression above in the form

$$\frac{\partial}{\partial s}\psi_{Y,X_1,\ldots,X_m}(s,t_1,\ldots,t_m)\bigg|_{s=0} =$$
$$\underbrace{\int\int\cdots\int}_{m-\text{times}} i(\sum_{j=1}^{m}\alpha_j x_j)\cdot\exp\left\{i\sum_{j=1}^{m}t_j x_j\right\}F_{\mathbf{X}}(d\mathbf{x}).$$

The necessary part of the Theorem follows from

$$\underbrace{\int\int\cdots\int}_{m-\text{times}} i(\sum_{j=1}^{m}\alpha_j x_j)\cdot\exp\left\{i\sum_{i=1}^{m}t_j x_j\right\}F_{\mathbf{X}}(d\mathbf{x})$$
$$= \underbrace{\int\int\cdots\int}_{m-\text{times}} \sum_{j=1}^{m}\alpha_j\frac{\partial}{\partial t_j}\exp\left\{i\sum_{j=1}^{m}t_j x_j\right\}F_{\mathbf{X}}(d\mathbf{x})$$
$$= \sum_{j=1}^{m}\alpha_j\underbrace{\int\int\cdots\int}_{m-\text{times}}\frac{\partial}{\partial t_j}\exp\left\{i\sum_{j=1}^{m}t_j x_j\right\}F_{\mathbf{X}}(d\mathbf{x})$$

and interchanging again differentiation and integration,

$$= \sum_{j=1}^{m}\alpha_j\frac{\partial}{\partial t_j}\underbrace{\int\int\cdots\int}_{m-\text{times}}\exp\left\{i\sum_{j=1}^{m}t_j x_j\right\}F_{\mathbf{X}}(d\mathbf{x})$$
$$= \sum_{j=1}^{m}\alpha_j\frac{\partial}{\partial t_j}\psi_{Y,X_1,\ldots,X_m}(0,t_1,\ldots,t_m).$$

The sufficiency part is easy to establish.

PROBLEM 4.2.6. Let $\{Y_k : k \in \mathbf{Z}_+\}$ be a sequence of i.i.d. random variables. Let T_p be a geometric random variable with parameter $p > 0$ and stochastically independent of $\{Y_k : k \in \mathbf{Z}_+\}$. Define the new random variable:

$$Z_p = \sum_{k=1}^{T_p} Y_k.$$

Prove that:

(a) if $E[Y_1] = a > 0$, then the distribution of the random variable $W_p = p \cdot Z_p$ converges weakly to an exponential distribution with parameter $1/a$ as $p \downarrow 0$;

(b) if $E[Y_1] = 0$ and $Var[Y_1] = \sigma^2 > 0$, then the distribution of the random variable $V_p = \sqrt{p} \cdot Z_p$ converges weakly to a random variable whose characteristic function is given by

$$\frac{1}{1 + \frac{1}{2}z^2\sigma^2}$$

as $p \downarrow 0$.

SOLUTION. Using the definition of characteristic function, the fact that the Y_k's are i.i.d. r.v.'s, and the assumption of independence between T_p and the sequence $\{Y_k : k \in \mathbf{Z}_+\}$, we have

$$\begin{aligned}
\psi_{W_p}(t) &= E[e^{itpZ_p}] = E[e^{itp\sum_{k=1}^{T_p} Y_k}] \\
&= E_{T_p}\left[E_{Y_1|T_p}[e^{itpY_1}]^{T_p} \mid T_p\right] = E_{T_p}\left[E_{Y_1}[e^{itpY_1}]^{T_p}\right] \\
&= E_{T_p}\left[E[e^{itpT_pY_1}]\right] = E_{T_p}[e^{T_p}E[e^{itpY_1}]] \\
&= E_{T_p}[e^{T_p}\psi_{Y_1}(tp)] = E_{T_p}[e^{T_p \log(\psi_{Y_1}(tp))}] \\
&= \phi_{T_p}(\log(\psi_{Y_1}(tp))
\end{aligned}$$

where $\phi_{T_p}(\cdot)$ is the moment generating function for T_p. Since $T_p \sim$ geometric(p), $p \in (0,1)$ the analytic form of $\phi_{T_p}(\cdot)$ is known and it yields

$$\psi_{W_p}(t) = \frac{pe^{\log(\psi_{Y_1}(pt))}}{1 - (1-p)e^{\log(\psi_{Y_1}(tp))}} = \frac{p\psi_{Y_1}(pt)}{1 - (1-p)\psi_{Y_1}(pt)}.$$

When $p \downarrow 0$ the last ratio is a $\frac{0}{0}$ form and its limit can be determined using De L'Hospital's Rule. This gives

$$\begin{aligned}
\lim_{p\downarrow 0} \psi_{W_p}(t) &= \lim_{p\downarrow 0} \frac{\psi_{Y_1}(pt) + pt\psi'_{Y_1}(pt)}{\psi_{Y_1}(pt) - (1-p)t\psi'_{Y_1}(pt)} \\
&= \frac{1}{1 - t\psi'_{Y_1}(0)} = \frac{1}{1 - itE[Y_1]} = \frac{1}{1 - ita}
\end{aligned}$$

which proves the first statement of the problem.

A procedure similar to that used above in part (a) shows that

$$\psi_{V_p}(t) = \frac{p\psi_{Y_1}(\sqrt{p}t)}{1 - (1-p)\psi_{Y_1}(\sqrt{p}t)}.$$

Again, $\lim_{p\downarrow 0} \psi_{V_p}(t)$ is a $\frac{0}{0}$ form whose indetermination can be eliminated by applying De L'Hospital's Rule. Thus

$$\lim_{p\downarrow 0} \psi_{V_p}(t) = \lim_{p\downarrow 0} \frac{\psi_{Y_1}(\sqrt{p}t) + \sqrt{p}(t/2)\psi'_{Y_1}(\sqrt{p}t)}{\psi_{Y_1}(\sqrt{p}t) - (1-p)/(2\sqrt{p})t\psi'_{Y_1}(\sqrt{p}t)}.$$

The last expression contains another indeterminate form that can be eliminated by a simple application of De L'Hospital's Rule:

$$\lim_{p\downarrow 0} \frac{(1-p)\psi'_{Y_1}(\sqrt{p}t)}{\sqrt{p}} = \lim_{p\downarrow 0}(1-p)t\psi''(\sqrt{p}t) = t\psi''(0) = ti^2E[Y_1^2] = -t\sigma^2.$$

Replacing this last expression in the previous limit for $\psi_{V_p}(t)$ we find

$$\lim_{p\downarrow 0} \psi_{V_p}(t) = \frac{1}{1+\frac{t^2\sigma^2}{2}}.$$

PROBLEM 4.2.7. Let Y be a random variable on some probability space such that its characteristic function, $\psi_Y(\cdot)$, is integrable. Let X be another random variable independent of Y having characteristic function $\psi_X(\cdot)$, and for γ a finite real constant define

$$Z = X + \gamma Y.$$

Let also $\alpha < \beta$ be two finite real numbers. Prove that

$$Pr[\alpha < Z < \beta] = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{e^{-i\beta t} - e^{-i\alpha t}}{-it}\psi_X(t)\psi(\gamma t)\,dt.$$

SOLUTION. By assumption, $\psi_Y(\cdot)$ is integrable and, using the independence between X and Y, we have also

$$\psi_Z(t) = \psi_X(t)\cdot\psi_Y(\gamma t).$$

This yields

$$|\psi_Z(t)| = |\psi_X(t)|\cdot|\psi_Y(\gamma t)| \leq |\psi_Y(\gamma t)|$$

and therefore the characteristic function $\psi_Z(\cdot)$ is also integrable[2]. This makes it possible to use the inversion theorem to compute the density function for Z :

$$f_Z(z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itz}\psi_Z(t)\,dt = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itz}\psi_X(t)\psi_Y(\gamma t)\,dt.$$

Then,

$$\begin{aligned}
Pr[\alpha < Z < \beta] &= \int_\alpha^\beta f_Z(z)\,dz \\
&= \int_\alpha^\beta \left(\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itz}\psi_X(t)\psi_Y(\gamma t)\,dt\right)dx \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \psi_X(t)\psi_Y(\gamma t)\left(\int_\alpha^\beta e^{-itz}\,dz\right)dt \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} \left(\frac{e^{-it\beta} - e^{-it\alpha}}{-it}\right)\psi_X(t)\psi_Y(\gamma t)\,dt.
\end{aligned}$$

[2]There exist r.v.'s whose ch.f.'s are not integrable over the real line. For example, $U \sim$ uniform$(-1,1)$ r.v. Its ch.f. is then given by $\psi_U(t) = \frac{e^{it}-e^{-it}}{2it}$ which is clearly not integrable over the whole real line.

PROBLEM 4.2.8. Let $\mathbf{x}, y$ be two $\mathbf{R}^p$ and $\mathbf{R}$-valued random variables defined on some probability space $(\Omega_{\mathbf{x}} \times \Omega_y, \mathcal{F}_{\mathbf{x}} \times \mathcal{F}_y, P)$. Let $f(\cdot)$ and $g(\cdot)$ be two different probability models for $(y, \mathbf{x})$. Prove that

$$f(y, \mathbf{x}) = g(y, \mathbf{x}) \text{ iff } f(y \mid \mathbf{a}^T\mathbf{x}) = g(y \mid \mathbf{a}^T\mathbf{x})$$

for all values of $\mathbf{a}^T\mathbf{x}$ in $\{\mathbf{a}^T\mathbf{x} : \mathbf{a} \in \mathbf{R}^p, \mathbf{x} \in \Omega_{\mathbf{x}}\}$.

SOLUTION. The $\Rightarrow$ implication is easy to prove as, in fact, from $f(y, \mathbf{x}) = g(y, \mathbf{x})$ follows that $f_{\mathbf{x}}(\mathbf{x}) = g_{\mathbf{x}}(\mathbf{x})$ and $f(y, \mathbf{a}^T\mathbf{x}) = g(y, \mathbf{a}^T\mathbf{x})$. From these equalities it follows in turn that $f(\mathbf{a}^T\mathbf{x}) = g(\mathbf{a}^T\mathbf{x})$ for any linear combination $\mathbf{a}^T\mathbf{x}$. Then,

$$f(y \mid \mathbf{a}^T\mathbf{x}) = \frac{f(y, \mathbf{a}^T\mathbf{x})}{f(\mathbf{a}^T\mathbf{x})} = \frac{g(y, \mathbf{a}^T\mathbf{x})}{g(\mathbf{a}^T\mathbf{x})} = g(y \mid \mathbf{a}^T\mathbf{x}).$$

The $\Leftarrow$ implication can be proved using characteristic fucntions. To this purpose, if we assume that $f(y \mid \mathbf{a}^T\mathbf{x}) = g(y \mid \mathbf{a}^T\mathbf{x})$ for all linear combination in $\{\mathbf{a}^T\mathbf{x} : \mathbf{a} \in \mathbf{R}^p, \mathbf{x} \in \Omega_{\mathbf{x}}\}$ then, it must be $\psi^{(f)}_{y|\mathbf{a}^T\mathbf{x}}(t) = \psi^{(g)}_{y|\mathbf{a}^T\mathbf{x}}(t)$ or, in other words:

$$E^{(f)}[e^{ity} \mid \sigma(\mathbf{a}^T\mathbf{x} : \mathbf{a} \in \mathbf{R}^p, \mathbf{x} \in \Omega_{\mathbf{x}})] = E^{(g)}[e^{ity} \mid \sigma(\mathbf{a}^T\mathbf{x} : \mathbf{a} \in \mathbf{R}^p, \mathbf{x} \in \Omega_{\mathbf{x}})].$$

The symbols $E^{(f)}$ and $E^{(g)}$ explain that the expectation is taken with respect to either model f or model g. Now,

$$\begin{aligned}
\psi^{(f)}_{y,\mathbf{x}}(t, \mathbf{s}) &= E^{(f)}[e^{ity+i\mathbf{s}^T\mathbf{x}}] \\
&= E[E^{(f)}[e^{ity+i\mathbf{s}^T\mathbf{x}} \mid \mathbf{a}^T\mathbf{x}]] \\
&= E[e^{i\mathbf{s}^T\mathbf{x}} E^{(f)}[e^{ity} \mid \sigma(\mathbf{a}^T\mathbf{x} : \mathbf{a} \in \mathbf{R}^p, \mathbf{x} \in \Omega_{\mathbf{x}})] \\
&= E[e^{i\mathbf{s}^T\mathbf{x}} E^{(g)}[e^{ity} \mid \sigma(\mathbf{a}^T\mathbf{x} : \mathbf{a} \in \mathbf{R}^p, \mathbf{x} \in \Omega_{\mathbf{x}})] \\
&= E^{(g)}[e^{ity+i\mathbf{s}^T\mathbf{x}}] = \psi^{(g)}_{y,\mathbf{x}}(t, \mathbf{s}).
\end{aligned}$$

The uniqueness of characteristic functions yields $f(y, \mathbf{x}) = g(y, \mathbf{x})$.

PROBLEM 4.2.9. Let $X_1, X_2, \ldots, X_n$ be i.i.d uniform(0, 1) random variables on some probability space $(\Omega, \mathcal{F}, P)$. Let $Y = \sum_{i=1}^n X_i$. Find the probability density function for Y.

SOLUTION. It is easy to show that the Laplace transform for X_1 is given by

$$\phi_{X_i}(t) = \frac{1 - e^{-t}}{t}$$

and, therefore, the Laplace transform for Y is given by

$$\phi_Y(t) = \sum_{k=0}^{n} (-1)^k \binom{n}{k} \frac{e^{-tk}}{t^n}.$$

In order to find the probability density for Y we notice that

(1) t^{-n} is the transform corresponding to $G(x) = \frac{x^n}{n!}$;

(2) $\frac{e^{-kt}}{t^n}$ is the transform corresponding to $H_k(x) = \frac{(x-k)_+^n}{n!}$

where x_+ denotes the function that equals x when $x \geq 0$ or $-x$ when $x < 0$. $G(\cdot)$ and $H_k(\cdot)$ are not probability measures as one can easily verify. However, Laplace transforms are defined for defective and proper probability measures alike. This means that the cdf for Y is given by

$$F_Y(y) = \frac{1}{n!}\sum_{k=0}^{n}(-1)^k\binom{n}{k}(y-k)_+^n.$$

Then

$$f_Y(y) = \frac{1}{(n-1)!}\sum_{k=0}^{n}(-1)^k(y-k)_+^{n-1}.$$